

Supplementary Materials of “Bridging from Goals to Tasks with Design Study Analysis Reports”, InfoVis 2017

Heidi Lam, Melanie Tory, and Tamara Munzner

[1 Design study papers considered](#)

[2 Summary of analysis goal assignments](#)

[2.1 Single-population analyses](#)

[2.2 Multiple-population analyses](#)

[3 Details of analysis goal and step assignments by publication](#)

[3.1 ABySS-E Section 6.2](#)

[3.2 BallotMaps Section 3](#)

[3.3 BallotMaps Section 4](#)

[3.4 BirdVis 6.1](#)

[3.5 BirdVis 6.2](#)

[3.6 BoxFish Section 5.2](#)

[3.7 DAVIEWER Section 6.4.1](#)

[3.8 DAVIEWER Section 6.4.2](#)

[3.9 Entourage Section 8.1](#)

[3.10 MovExp Section 5.3](#)

[3.11 MovExp Section 5.4](#)

[3.12 MulteeSum Section 7.1](#)

[3.13 MulteeSum Section 7.2](#)

[3.14 MulteeSum Section 7.3](#)

[3.15 NeuroLines Section 9.1](#)

[3.16 NeuroLines Section 9.2](#)

[3.17 Paramorama Section 5.1](#)

[3.18 Poemage Section 8.1](#)

[3.19 Ravel Section 5.1](#)

[3.20 Ravel Section 5.2](#)

[3.21 SellTrend Section 7](#)

[3.22 SignalLens Section 5.1](#)

[3.23 SignalLens Section 5.2](#)

[3.24 SnapShot Section 5.1](#)

[3.25 SnapShot Section 5.2](#)

[3.26 SnapShot Section 5.3](#)

[3.27 SoccerStories Section 6.1](#)

[3.28 TenniVis Section 5.2](#)

[3.29 TenniVis Section 5.3](#)

[3.30 Vials Section 6.1](#)

[3.31 Vials Section 6.2](#)

[3.32 Weaver Section 8.1](#)

[5 Excerpt from the video analysis in which we applied our framework](#)

[6 Excerpt from an analysis log file in which we applied our framework](#)

1 Design study papers considered

Code Name	Title	First Author	Year
ABySS-E	ABySS-Explorer: Visualizing Genome Sequence Assemblies	Nielsen	2009
BallotMaps	BallotMaps: Detecting Name Bias in Alphabetically Ordered Ballot Papers	Wood	2011
BirdVis	BirdVis: Visualizing and Understanding Bird Populations	Ferreria	2011
BoxFish	Visualizing Network Traffic to Understand the Performance of Massively Parallel Simulations	Landge	2012
DAViewer	Facilitating Discourse Analysis with Interactive Visualization	Zhao	2012
Entourage	Entourage: Visualizing relationships between biological pathways using contextual subsets	Lex	2013
MovExp	MovExp: A Versatile Visualization Tool for Human-Computer Interaction Studies with 3D Performance and Biomechanical Data	Palmas	2014
MulteeSum	MulteeSum: A Tool for Comparative Spatial and Temporal Gene Expression Data	Meyer	2010
NeuroLines	NeuroLines: A Subway Map Metaphor for Visualizing Nanoscale Neuronal Connectivity	Al-Awami	2014
Paramorama	Visualization of Parameter Space for Image Analysis	Pretorius	2011
Poemage	Poemage: Visualizing the Sonic Topology of a Poem	McCurdy	2016
Ravel	Combing the Communication Hairball: Visualizing Large-Scale Parallel Execution Traces using Logical Time	Isaacs	2014
SellTrend	SellTrend: Inter-Attribute Visual Analysis of Temporal Transaction Data	Liu	2009
SignalLens	SignalLens: Focus+Context Applied to Electronic Time Series	Kincaid	2010
SnapShot	SnapShot: Visualization to Propel Ice Hockey Analytics	Pileggi	2012
SoccerStories	SoccerStories: A Kick-off for Visual Soccer Analysis	Perin	2013
TenniVis	TenniVis: Visualization for Tennis Match Analysis	Polk	2014
VariantView	Variant View: Visualizing Sequence Variants in their Gene Context	Ferstay	2013
Vials	Vials: Visualizing Alternative Splicing of Genes	Strobelt	2016
Weaver	Visually Comparing Weather Features in Forecasts	Quinan	2016

2 Summary of analysis goal assignments

Section numbers refer to sections in the design study papers where the analysis reports were found. We indicate multiple instances of the analysis goal in the same paper section by the count in parentheses, e.g., BallotMaps §3 (2)

2.1 Single-population analyses

Explore	Describe	Explain	Confirm
<p>Discover Observation¹</p> <p>Poemage §8.1 SignalLens §5.2</p>	<p>Describe Observation (Item)</p> <p>NeuroLines §9.1 SignalLens §5.1</p> <p>Describe Observation (Aggregate)²</p> <p>BallotMaps §3</p>	<p>Identify Main Cause (Item)</p> <p>AbySS-E §6.2 DAViewer §6.4.1 Entourage §8.1 Paramorama §5.1 Ravel §5.1 Ravel §5.2 SnapShot §5.1 TenniVis §5.2 TenniVis §5.3</p> <p>Identify Main Cause (Aggregate)</p> <p>SellTrend §7</p>	<p>Collect Evidence</p> <p>ABYSS-E §6.2 MulteeSum §7.2 Ravel §5.1 (2) Ravel §5.2 (2) TenniVis §5.2 Vials §6.2 Weaver §8.1</p>

¹ In addition, Discover Observation is the first step of many analysis reports, which carried on as other goals:

AbySS-E §6.2	Paramorama §5.1	SnapShot §5.1
BallotMaps §3 (2)	Ravel §5.1	SoccerStories §6.1 (2)
BallotMaps §4	Ravel §5.2	TenniVis §5.2
DAViewer §6.4.1	SellTrend §7	Vials §6.2
MulteeSum §7.2	SignalLens §5.1	

² In addition, Describe Observation is a precursor to multiple population analyses as it outputs population definitions, so this analysis goal is the first step in the following reports:

BallotMaps §4	Entourage §8.1 (2)	Ravel §5.2
BirdVis §6.1	MovExp §5.3	SnapShot §5.1
BirdVis §6.2	MovExp §5.4	SoccerStories §6.1
BoxFish §5.2	MulteeSum §7.1	TenniVis §5.2
DAViewer §6.4.1	MulteeSum §7.3	TenniVis §5.3
DAViewer §6.4.2	NeuroLine §9.2	Vials §6.1 (2)

2.2 Multiple-population analyses

Explore	Describe	Explain	Confirm
<None>	Compare Entity BirdVis §6.1 BirdVis §6.2 MulteeSum §7.1 MulteeSum §7.3 Entourage §8.1	Explain Differences BallotMaps §4 BoxFish §5.2 DAViewer §6.4.1 DAViewer §6.4.2 MovExp §5.4 SoccerStories §6.1 TenniVis §5.3 Vials §6.1 Weaver §8.1	Evaluate Hypothesis BoxFish §5.2 MovExp §5.3 NeuroLines §9.2 SnapShot §5.2 SnapShot §5.3 SoccerStories §6.1 Vials §6.1

3 Details of analysis goal and step assignments by publication

For each paper identified in the table “[Design Study Papers Considered](#)” above, we open-coded the analysis reports (sections where the paper reported a target user performing analysis), where we identified the steps in the analysis and the analysis goals. The following tables show the codes assigned to each relevant paper section.

3.1 ABySS-E Section 6.2

Goal	Step	Paper Text
Discover Observation	Note observation	By imposing these contig alignment annotations onto the assembly structure, we clarify where the assembly order and the alignment order agree. For example, the strings of uninterrupted orange or blue contigs at the bottom of the graph indicate agreement. Inconsistencies are immediately apparent as interconnections between different colored contigs, such as in the center of the graph.
Identify Main Cause (Item)	Identify likely dominant cause	It is interesting to note that many of these connection points are made of short contigs, suggesting that repetitive elements or sequence errors are the cause of such ambiguities.
	Focus on instance	The inversion breakpoint itself is resolved quite clearly, with the black breakpoint contig flanked by a string of blue contigs on one side and orange contigs on the other (Figure 11).
	Identify likely dominant cause	The adjacency of dark orange and blue contigs highlights the nature of the inversion event (compare Figure 10b).
	Focus on instance	Closer inspection of this region reveals a lighter blue contig close to the breakpoint,
	Identify likely dominant cause	highlighting an inconsistency between the alignment order (represented by the color gradient) and the assembled order.
Collect Evidence	Form hypothesis	Analyst must now decide which ordering is correct. Is there an error in the map of the inversion event (Figure 10b) which was deduced using other experimental methods, or is there a mis-assembly?
	Identify evidence to support hypothesis	Interactive exploration of the paired read information (Figure 12) reveals that the large suspect contig, now colored dark gray, connects only to smaller contigs mapping to regions further from

		breakpoint. Because the read pair information supports the alignment order, we suspect a spurious mis-assembly. Interactive data exploration greatly facilitates this kind of analysis.
	Assess hypothesis	The detailed view of the breakpoint region (Figure 11) also offers a guide as to where small ambiguously aligned contigs (gray) should be placed. Again, paired read information can be used to make judgements about which contigs belong and which ones may be spurious alignments.

3.2 BallotMaps Section 3

Goal	Step	Paper Text
Discover Observation	Note observation	This particular depiction shows some evidence of name bias that itself is geographically and party related. If electoral success were based only on party preference and candidate suitability there should be no relation with ballot paper position. We would therefore expect the horizontal length of each dark bar to be roughly similar for each party in each borough; any variation being random.
Describe Observation (Aggregate)	Identify attribute(s) to define/refine population(s)	This is particularly evident in the boroughs of Islington, Richmond upon Thames, Sutton and Lewisham where candidates listed first in their party are more likely to be elected than those second or third.
	Identify attribute(s) to define/refine population(s)	Some boroughs show this effect largely for certain parties, such as the (blue) Conservatives in Ealing and the (orange) Liberal Democrats in Brent.
	Identify exception to observation	A few boroughs appear to show no ordering effect, such as Bromley and Croydon and a few others where party preference dominates the distribution of elected councillors, such as Newham and Barking and Dagenham.
	Identify exception to observation	the prosperous boroughs west and southwest of central London show a strong ordering effect for right of center Conservative candidates where there may be more tendency for voters to split their three votes between Conservative and Liberal Democrat candidates.
	Describe population	on average, a candidate listed first in their party is 6.3 times as likely to get the most votes in their party than a candidate listed third. The effect is strongest for Liberal Democrat candidates; a candidate listed first in their party is 8.6 times more likely to get the most party votes than one listed third.
Discover Observation	Note observation	Assuming an expected value for each candidate of exactly one third of the total votes for their party in their ward, the chi ballotMap clearly shows the systematic ordering effect when

		values sorted graphically from top to bottom by order within party then order on ballot paper. If there was no ordering effect, we would expect a random distribution of purple and green cells.
Describe Observation (Aggregate)	Identify attribute(s) to define/refine population(s)	By breaking down the distributions by party, it is also evident that the strongest ordering effect is for Liberal Democrat candidates (the top and bottom thirds of the central column in Fig. 6 are generally darker green and darker purple than the top and bottom thirds of the left and right columns representing the other two parties). Labour candidates positioned first in their party show a slightly stronger ordering effect than Conservative candidates.

3.3 BallotMaps Section 4

Goal	Step	Paper Text
Discover Observation	Note observation	<p>Interactive query of anomalous candidate names suggested there might be an association with the apparent ethnicity implied by the name. Initially we created tag clouds comprising all names of candidates positioned first in their party with a negative residual value - that is, those who received less than the average percentage party vote for candidates positioned first.</p> <p>In order to indicate whether this distribution of names was systematically different to those of all candidates in alpha1 position, we then compared this distribution with tag clouds of random selections drawn from the alpha1 sample. We borrowed from the process of graphical inference [26] to compare the observed values (alpha1 names with negative residuals) with a null hypothesis assuming no structure to anomalies (random samples from alpha1). While this indicated there might be some degree of ethnicity bias present, we wished to examine the structure of that bias in more detail.</p>
Describe Observation (Aggregate)	Identify attribute(s) to define/refine population(s)	To investigate possible ethnicity bias, we allocated each candidate to a class relating to the likely ethnic origin of their name using OnoMAP [12, 16]. This classification, evaluated for use in public health policy [10], compares given and family names to classify each pair into one of 16 possible OnoMAP ethnic groups. The numbers of candidates in some of the OnoMAP groups were too small to draw significant conclusions, and there was also a question of the discriminating power of voters in being willing or able to distinguish between certain groups. We therefore chose to group all candidates into two broad groups – ‘English or Celtic’, comprising the OnoMAP ‘English’ and OnoMAP ‘Celtic’ groups of names that are likely to originate in the British Isles, and ‘Other Name Origins’, comprising all other name origin groups.
Explain Differences	Identify attribute difference(s) between populations	Fig. 7 shows the chi values for all candidates broken down by these two super-groups. The BallotMap shows that there are approximately similar numbers of candidates from both ethnic super-groups in all alpha positions, but that name ordering bias is much higher in the ‘English or Celtic’ group. The ‘other name

		origins' group shows many more purple candidates with fewer votes than expected in the alpha1 position. This suggests that, for some candidates at least, a propensity not to select a candidate due to their non British Isles name origin outweighs a propensity to select them because they are positioned first within their party on the ballot paper.
	Identify attribute difference(s) between populations	To explore whether this effect had any geographical component, we constructed BallotMaps showing the chi values by ethnic super-group for each borough (see Fig. 8). The ballotMap shows that the ethnicity of candidates varies by geography, for example the western boroughs of Harrow, Brent, Ealing and Hounslow having a higher proportion of 'other name origin' candidates compared with southern boroughs of Richmond, Merton, Sutton, Bromley and Greenwich. All boroughs show a name bias in the 'English or Celtic' supergroup (upper thirds greener than lower thirds), but in the 'other name origins' group, the pattern is more varied. In many of the outer boroughs, the alpha1 candidates show fewer than expected votes (purple cells in the top left of the 32 squares representing each borough), for example Brent, Harrow, Kingston, Sutton, Bromley and Greenwich. In contrast, some of the inner boroughs with higher numbers of candidates in the 'other name origins' super-group show a name ordering bias within this group that is similar to or stronger than that seen in the 'English or Celtic' group (e.g. Hackney, Tower Hamlets, Newham).

3.4 BirdVis 6.1

Goal	Step	Paper Text
Describe Observation (Aggregate)	Identify attribute(s) to define/refine population(s)	The Palm Warbler is especially interesting in that it has two populations. A subspecies of eastern breeders (<i>D. p. hypochrysea</i> ; Yellow Palm Warbler) nests in Atlantic Canada and winters from northern Florida to east Texas, while another (<i>D. p. palmarum</i> ; Western Palm Warbler) nests further west in Canada but winters further east, from the southeastern U.S. to south Florida and the western Caribbean
Compare Entities	Identify attribute similarities and differences between populations	The scientists produced occurrence maps and occurrence variation maps to visualize the Palm Warbler spring migration. In Figure 9, we can see that, on April 13 (first column), the migration starts: the southeastern wintering range shows well on the top occurrence map, with peak occurrence in Florida; declines in Florida and increases in New England indicate that migration is underway with Yellow Palms leaving the southeast and heading to Atlantic Canada. On April 27, the migration of Yellow Palm is already on the decline, with lower occurrence in New England as compared to the previous week, but an obvious push to the north-northwest as Western Palms move from Florida towards the Great Lakes region. Subsequent figures show a clear picture of northward passage, with Yellow Palm migration ending well before that of the western population. By May 18, Palm Warblers are less common

		everywhere than they were the previous week, indicating that spring migration for the species as a whole is drawing to a close.
	Relate findings to domain	This visualization of bird occurrence patterns, showing multiple dates and both the occurrence and occurrence variation, has provided new insights and hypotheses. Past authors have discussed the migration of the Yellow Palm Warbler as heading northeastward up the Atlantic Coast (e.g., [8]). However, Figure 9 (on April 27 and May 4) shows high Palm Warbler occurrence New England and the Southeastern U.S., but the species never seems to reach high occurrence in the mid-Atlantic states, which shows up as a wedge of low occurrence on Figure 9, April 27. The pattern actually suggests that a significant proportion of the population may shortcut across the Atlantic Ocean. While intentional autumn movement over the western Atlantic Ocean is well-known in several species of shorebirds and some passerines such as Blackpoll Warbler [27, 28], it has not been suggested to be a significant migration path in spring. STEM models do suggest that this may occur in spring in both Yellow Palm and Blackpoll Warblers, and anecdotal observations from field birders (A. Farnsworth, M. Iliff pers. obs.) provide further support for the possibility of an offshore migration path in spring.

3.5 BirdVis 6.2

Goal	Step	Paper Text
Describe Observation (Aggregate)	Identify attribute(s) to define/refine population(s)	In addition to map views, the scientists used the tag cloud lenses to visualize how the habitat preferences change on the maps. As Figure 1 illustrates, by creating multiple tag cloud lenses for different regions and examining different dates, they were able to identify interesting regions where the local importance of the predictors differs from the importance computed over the entire map. Figure 11 highlights three relationships: Indigo Bunting distribution, date, and the important of habitat (NLCD classes) predictors in relation to predicted species occurrence.
Compare Entities	Identify attribute similarities and differences between populations	On the left, we see the species' core breeding range (dark orange), with the most important factor affecting their predicted occurrence being an association with Deciduous Forest. But during fall migration (shown on the right), we visualize a different story. The core population has shifted southward (dark orange), and now the most powerful association for predicted occurrence is Cultivated Crops, with Deciduous Forest and Pasture weighing heavily in the mix.
	Relate findings to domain	This apparent habitat shift from shrubby thickets to more open, grassy, agricultural areas fits well with the species known biology [31]. During the breeding season, Indigo Buntings require protein-rich insects to raising young. But after breeding, the buntings switch to seeds that allow them to gain fat reserves necessary to fuel their journey over the Gulf of Mexico.

3.6 BoxFish Section 5.2

Goal	Step	Paper Text
Describe Observation (Aggregate)	Identify attribute(s) to define/refine population(s)	<p>As discussed above, the behavior of the default mapping is fairly predictable and our tool was mainly used to validate prior expectations. However, when experimenting with different mappings the network traffic becomes far less predictable. Through extensive experiments in which the aggregated bandwidth was recorded for different mappings (see Table 1) the performance experts knew a priori that certain mappings can achieve significantly better performance than the default. However, the causes of these differences were unclear.</p> <p>We experiment with five different mapping types each with different characteristics. In addition to the default mapping we use an XYZT mapping, a tiled layout, and two different tilted layouts, one tilted just along the z-axis and one tilted along both z and y. Fig. 7 shows the node mappings using the 3D view for an $8 \times 8 \times 8$ hardware torus using 16 slabs of 16×8 patches. The XYZT layout spreads out the x-communicators by distributing them on individual nodes rather than four to a node as the default. As a result a single slab is spread between two planes of the torus rather than the half plane of the default. The tile mapping is similar to the XYZT mapping, but changes the orientation of the layout by mapping slabs into tiles perpendicular to slabs in the default mapping. The tiltZ mapping starts from the tiled layout but then “tilts” each yz-plane of the torus in the z direction. This drastically increases the size of their bounding box. TiltZY further modifies tiltZ by introducing a second tilt in the y direction. This increases the bounding box of the nodes even further. In particular, these may have been effects particular to a specific number of nodes or configuration which would raise doubts about the scalability of the mappings. Furthermore, understanding the difference in network traffic between various mappings in detail can provide insights into the design of even better mappings.</p>
Explain Differences	Identify attribute difference(s) between populations	<p>Given this context, we compare network traffic for the five different node mappings at different scales. These are concisely described by the minimaps of the 2D projection, as shown in Figs. 8(a) (512 nodes) and 8(b) (1,024 nodes). By using an interactive slider that controls all five views simultaneously we can provide potential explanations for the performance measurements in Table 1. In particular, the TXYZ mapping entirely excludes communication in the z-direction, strongly clustering communication in the other two as only half planes communicate. Instead, the XYZT mapping spreads out the nodes more and utilizes some z-links within a slab.</p>
	Relate findings to domain	<p>The minimaps clearly show a more even distribution of communication load even though the same patterns as the TXYZ mappings are apparent. This provides a significant boost in</p>

		performance by more than doubling the total bandwidth.
	Identify attribute(s) to define/refine population(s)	The tile mapping acts roughly like a rotated XYZT mapping and shows very similar behavior and performance.
	Identify attribute difference(s) between populations	The TiltZ mapping however further balances the communication. In particular, note how both the top and bottom minimap indicate (relative) increases in x communication. Since the total amount of communication is independent of the mapping, this increase actually indicates a further balancing of the communication.
	Relate findings to domain	These experiments strongly suggest that evenly distributing the traffic leads to better bandwidth usage. Part of the current hypothesis is that increasing the effective bounding box sizes of the slabs and evening out their aspect ratios drastically increases the number of potential routes a packet can choose. Coupled with the dynamic routing of the BG/P system, this may be the cause for the increase in aggregate bandwidth. Note that the fact that providing more routes increases the potential bandwidth is expected. However, under the current thinking pF3D is not bandwidth limited thus the fact that increasing the available bandwidth caused increased traffic rates was a novel finding. Additionally, the best mappings clearly increase the distance packets must travel which, however, does not seem to have a negative effect. This is especially surprising for the default mapping as much of its x communication is restricted to intra-node communication which is expected to be significantly faster than any inter-node messaging.

3.7 DAVIEWER Section 6.4.1

Goal	Step	Paper Text
Describe Observation (Aggregate)	Identify attribute(s) to define/refine population(s)	In this scenario, the user wants to investigate the flaws of the HILDA parser, a popular algorithm in the domain. She starts by glancing at the overview which presents the similarity scores, and finds that overall, the parser is performing fairly well. She identifies the two documents with the highest (0.86 and 0.83) and lowest (0.52 and 0.54) scores and selects them for deeper analysis. The gold standard, and the HILDA output for these four documents are loaded in the detail panel.
Explain Differences	Identify attribute difference(s) between populations	The user immediately finds out that the documents causing errors are much longer than the other ones, which is reasonable, as typically, the more content, the more challenging the parsing. Likewise, the discourse trees are large and difficult to read as a whole structure.
Discover Observation	Note observation	In order to get an overall idea of where the algorithm fails, the user reduces the trees to their compact representations to observe the distribution of scores, groups and so forth. From the vertical

		compact view, she observes that while the distribution of nodes into groups is similar to the gold standard, the scores of HILDA are very low (Figure 8a). Next she expands the tree views, and at the same time compacts the short documents since they are not the focus a the moment.
Identify Main Cause (Item)	Focus on instance	With the help of the heatmap background of the dendrogram, the user identifies where the error first occurs:
	Assess hypothesis based on external information	HILDA groups EDUs 16 and 17 as early as the second level whereas the gold standard keeps the branches separated up to level 17 (Figure 8b).
	Identify likely dominant cause	Thus she found that this first error, which propagates to the root node, is a major problem that strongly affects the overall parsing.
	Focus on instance	By looking at the text, she finds that the EDU 17 says “individual prosperity inevitably would result” where the keyword “result” is a critical indicator of the Cause relation. Yet, the HILDA parser groups this node under an Elaboration relation.
	Examine related data to understand observation	To get some context, the user switches to the continuous text display to comfortably read the sentence with the problematic EDU.
	Assess hypothesis based on external information	She finds that EDUs 7-16 as a whole are the summary of previous content, that should be grouped together in a branch under the Cause relation, with EDU 17, as indicated by the gold standard. The user thus selects the group of nodes and comments on her finding on the annotation panel.
	Examine finding(s) with other instances of observation	Meanwhile, she wonders if such errors happen elsewhere. She adds more trees with low scores to the detail panel, and through the query panel, looks for other Causal relation structures with branches containing the keywords “because” and “as a result”. A close examination of the results reveals that HILDA incorrectly groups the nodes or mislabels the relation (Elaboration or Explanation) and adds notes each time she finds such error in the dataset for further consultation. Indeed, the above findings provide hints for improving parsers, by taking more careful consideration of the content under a Cause relation

3.8 DAVIEWER Section 6.4.2

Goal	Step	Paper Text
Describe Observation (Aggregate)	Identify attribute(s) to define/refine population(s)	After identifying several issues in the HILDA parser, the user wants to investigate if and how tuning different parameters affects the outputs of her own parsers. She appends the result of three variations of her algorithm to the overview matrix (referred to as algorithms A1, A2 and A3). In this scenario, she sets HILDA as the reference column, since she wants to compare where the algorithms differ in performance. Adopting a similar approach as that of the previous scenario, she first glances at the overview and finds out that the fourth column (A2, corresponding to the condition

		“no N-grams feature”) provides the most differing results, and that the rightmost column (A3, condition “no syntactic prefix and suffix”) provides a very similar parsing as that of HILDA (Figure 9a). She selects a subset of four rows (two documents with the most similar and most differing results) and three columns (HILDA, A2 and A3) for further analysis in the detail panel.
Explain Differences	Identify attribute difference(s) between populations	By looking at the different tree representations (i.e., compact views, dendrogram and icicle), the user discovers that the trees generated by A2 usually contain many more levels and are more skewed, indicating that the classifier cannot find clear grouping pivots. To make the differences more visible, the user decides to fade out the branches similar to those of HILDA by filtering them out through the use of the interactive similarity score legend, and switches to the icicle view to analyze the relation types.
	Identify attribute difference(s) between populations	She observes that most of the remaining nodes are labeled with Elaboration or Same Unit by A2, which is unsurprising, since the two relations are the most common ones in that specific corpus. After deactivating the latter in the relation legend, she clearly observes that A2 hardly identifies any other relations, indicating that the N-grams feature, which was deactivated in A2, is essential for the relation classifier.
	Identify attribute difference(s) between populations	Our user wants next to compare the performance of a particular relation: Contrast. She looks for phrases indicative of this relation: “but,” “on the contrast,” etc., coupled with a structure-based query around the Contrast relation. She observes that A2 fails at identifying such relations, while it is usually well labeled by discourse parsers (Figure 9b). She further comments on her findings by adding a note that considering a certain number of EDUs as a whole (as N-grams does) is a good parser feature for identifying the real rhetorical relations.

3.9 Entourage Section 8.1

Goal	Step	Paper Text
Describe Observation (Aggregate)	Identify attribute(s) to define/refine population(s)	Due to its immediate relevance, the expert started by loading the ErbB signaling pathway into Entourage. By searching for related pathways she found several cancer-specific pathway maps. The pathways Glioma and Non-small cell lung cancer ranked among the top on the list (see Figure 7). The pathways Glioma and Non-small cell lung cancer ranked among the top on the list (see Figure 7). She commented that this indicates that the ErbB signaling pathway is a key player in these diseases. For the ErbB pathway map, our collaboration partner was interested in the experimental data for the genes in the path that leads from ErbB receptors to Myc, a gene known to regulate cell growth. She also noticed that ErbB2 was highlighted with a red exclamation mark indicating high variance in the copy number data. She thus

		selected the genes of this path for an in-depth analysis. She then looked at this path's gene expression data in the embedded enRoute view and combined it with sensitivities to Erlotinib and Lapatinib. For the analysis, cell lines were grouped by their tissue of origin (e.g., breast, ovary, liver, etc.) and sorted by sensitivity to Lapatinib.
Compare Entities	Identify attribute similarities between the populations	Her first observation, when looking at the experimental data, was that the two drugs displayed inhibitory activities across cell lines from many different tissues.
	Identify attribute differences between the populations	The cell lines from lung, breast and three other tissues were in general most responsive.
	Identify attribute similarities between the populations	The set of cell lines that were responsive to Erlotinib and Lapatinib largely overlapped,
	Identify attribute differences between the populations	although Lapatinib showed a broader spectrum of activity than Erlotinib.
	Identify differences between the populations	She found a strong co-occurrence between ErbB2 mRNA over-expression and sensitivity to Lapatinib in lung and breast cancer cell lines, a trend that was less apparent or not observed at all for other responsive cell lines.
	Identify attribute(s) to define population(s)	She then chose to focus on cell lines from breast and also investigated copy number variation for these cell lines
	Identify attribute similarities between the populations	For most breast cancer cell lines that over-expressed ErbB2, high copy numbers of this gene were found, i.e., the increased expression could generally be traced back to an increased copy number.
	Identify differences between the populations	Interestingly, only two breast cancer cell lines that showed strong over-expression of ErbB2 did not respond to Lapatinib treatment.
Describe Observation	Identify attribute(s) to	She then tried to find the cause for this effect and examined the expression of downstream genes in the pathway.

(Aggregate)	define/refine population(s)	
Identify Main Cause (Item)	Identify likely dominant cause	She found that for these two cell lines, the gene Ras was strongly under-expressed (also shown in Figure 7).
	Examine finding(s) with other instances of observation	Based on this observation she started to investigate whether other cancer-related pathways contain the same signaling cascade, i.e., path. She selected Ras as the focus node of her analysis, which revealed several other pathways that contain the same ErbB signaling cascade.

3.10 MovExp Section 5.3

Goal	Step	Paper Text
Describe Observation (Aggregate)	Identify attribute(s) to define/refine population(s)	<p>The HCI experts were interested in identifying optimal input regions for interfaces controlled by arm movements. By optimality, they referred to input regions with highest performance and lowest ergonomic costs. One case was a public display. These interactive surfaces often require keeping the arm extended during interaction, i.e., the ergonomics are of great interest here. Their analysis was focused on horizontal input regions (strips). The data set came from recordings of a male subject selecting targets on the display. Figure 1 shows the setup.</p> <p>The circular directions visualization allowed selecting movements that correspond to horizontal input regions. The optimality constraint was brushed on a scatter plot showing muscle activation against throughput (Figure 1). The two selections were combined using the and operator of MovExp. The outcome was shown on a intuitive case-specific visualization where a photo of the public display and some circles were combined to show the input regions.</p>
	Evaluate Hypothesis	Form hypothesis
Evaluate Hypothesis	Identify attribute differences between populations	The optimal region was identified to be in the middle of the display. In this region, the mean movement inaccuracy is 17mm and the mean interaction throughput is 13.8bits/s. Also, the index of energy expenditure, which is activation of all muscles integrated over movement, is 145.6 abstract units. By contrast, the movements in the non-optimal region are 19.2% less accurate (mean inaccuracy 20.5mm) and as result the mean interaction throughput is 1.5 bits/s lower (12.3bits/s).
	Identify attribute differences between populations	The energy expenditure difference is more dramatic and shows 200.9 abstract units, which is 39% higher comparing to the optimal region.

	Relate findings to domain	In summary, the interaction in the optimal region is faster and less energy-demanding.
	Assess hypothesis	The result confirmed the experts' hypothesis that placing frequently used buttons on the top of the display is detrimental. The implication was to either lower the display or place the buttons in the middle part to the left and right sides of the content view.

3.11 MovExp Section 5.4

Goal	Step	Paper Text
Describe Observation (Aggregate)	Identify attribute(s) to define/refine population(s)	A recurring interest in empirical HCI is the comparison of alternative designs. Our HCI experts were interested in comparing different methods for controlling a plane in a flight simulator: 1) the bird, where arms are extended to the side, 2) the steering wheel, where arms are extended and rotate for control, and 3) the arm flexor, where the right arm is lowered and flexed. They wanted to identify the least fatiguing method. Due to the high number of muscles, comparing their respective activations using a line plot or a bar plot would be too slow and unintuitive. The muscle views immediately conveyed stark differences of the muscle activations. Figure 8 shows them rendered for each condition.
Explain Differences	Identify attribute differences between populations	The HCI experts identified the steering wheel as the best method. It recruited mostly the muscles of the lower body together with some postural muscles and the muscles of the neck. This includes the gluteus muscles as well as the gracilis, splenius capitis and levator scapularis. By contrast, the bird recruited the stronger muscles of the upper back, shoulder, chest and arm. This includes the deltoids, infraspinatus, pectoralis major, biceps, brachialis and serratus anterior. All the other muscles were recruited moderately in a similar way in both of the cases.
	Relate findings to domain	While the absolute difference between the two sums of activations of all muscles is quite small, the general effect of the differences is significant. In fact, the lower body and the postural muscles usually contain a higher percentage of fatigue-resistant fibers than the upper body muscles. This is beneficial for the steering wheel, where the lower body muscles do not get fatigued as fast as the upper body muscles in the bird case.

3.12 MulteeSum Section 7.1

Goal	Step	Paper Text
None	Context	While exploring an early version of the Dpse VE using our second prototype system discussed in Section 6, the biologists discovered that the data set was in fact plagued by significant low-level noise.

Describe Observation	Identify attribute(s) to define/refine population(s)	Using summaries that compared the Dmel and early Dpse data sets, they noticed that the expression profiles of many cells with high summary values, i.e., those cells with a large dissimilarity from their aggregation group cells,
Compare Entities	Identify attribute similarities between populations	were actually quite similar to the expression profiles of their aggregation group in terms of the most highly expressed genes.
	Identify attribute(s) to define/refine population(s)	By analyzing and comparing the expression profiles of many of these high-value cells,
	Describe population	they found that the summary values were being dominated by contributions from low-level noise in the Dpse data for a handful of genes. An example of this noisy data is shown for a Dpse aggregation group in Figure 4(a).
	Relate findings to domain	These high, but uninteresting, summary values were masking the types of expression profile differences the biologists were hoping to find. They made the decision to generate a new Dpse VE, the data for which is shown in 4(b) for the same set of cells.

3.13 MulteeSum Section 7.2

Goal	Step	Paper Text
Discover Observation	Note observation	Investigating the expression profile for the Dmel cell shown in Figure 4(b), this biologist noticed that the RMS metric did not match the cell with the biologically most similar cell from the aggregation group.
Collect Evidence	Form hypothesis	The RMS metric is sensitive to small variations over the entire expression profiles, at times obscuring the biologically significant differences in just a small set of genes
	Identify evidence to support hypothesis	In this case, the RMS best match, shown in the second row of the curvemmap, has significant differences in the expression levels of both the hkb and gt genes compared to the selected cell shown in the top row. This is in contrast to the other aggregation group cells, which match the selected cell's expression trends more closely for these genes. This mismatch indicates that the RMS metric is sensitive to small variations over the entire expression profiles, at times obscuring the biologically significant differences in just a small set of genes. For this example, a possible next computation would weight the contributions for specific genes of interest more heavily than others, or filter low-level noise.
	Identify evidence to support hypothesis	Another biologically meaningful trend that the RMS 50 summary obscures is shown in Figure 5, where the expression profile of the selected cell significantly varies from those in its aggregation group in just a single gene, prd. The summary value for this selected cell was only moderately high, but the potential biological implications for this variation are very important to this biologist. She noted that

		a significant variation in a single gene could indicate a novel regulation mechanism, and this specific example provides an interesting direction for follow-up experiments. Similar to the previous example, this observation indicates the need for computations that are sensitive to variations in just a single, or a small set of, genes in the expression profiles.
	Identify evidence to support hypothesis	In a third example shown in Figure 6, two cells with roughly similar summary values were selected by the biologist. One cell resides in the anterior of the embryo, while the other in the posterior. Of interest is the distribution of metric values for the aggregation groups — shown in Figure 6(a) are the values for the anterior cell and in Figure 6(b) are the values for the posterior cell. She noticed for the posterior cell the distribution has a very long, flat tail of low values, indicating a potentially large neighborhood of similar cells in the comparison embryo, compared with the relatively few similar cells for the anterior cell. This finding reiterated to the biologist that the RMS computation can assign similar summary values to cells that have very different comparison trends to their aggregation groups.

3.14 MulteeSum Section 7.3

Goal	Step	Paper Text
Describe Observation (Aggregate)	Identify attribute(s) to define/refine population(s)	She created a summary for Dmel from the expression level of the hb gene at a single time point at each cell in the VE. This summary is shown in Figure 7. Of particular interest is the band of high values in the posterior of the embryo. On either side of this band two groups of cells were created to analyze how the expression profiles change moving inwards towards the band. These groups are the left group and the right group, and both were generated by manually selecting cells in MulteeSum. The groups were loaded into the curvemap, with the left group assigned an orange color and the right group assigned a purple color. A high-value cell from the middle of the band is selected for comparison.
	Identify attribute similarities between populations	The expression profile of the selected cell shows consistently low levels of all regulator genes and high levels of hb.
Compare Entities	Identify attribute difference(s) between populations	In the left group, the repressor gene gt is high, while in the right group the repressor gene till is high — the biologist expected to see this from prior knowledge about these genes.
	Identify attribute(s) to define/refine population(s)	[Implicit; focusing on the right group and compared the expression of till and hb]
	Identify attribute	Several trends she was not aware of, however, were immediately

	difference(s) between populations	obvious to her. She noted that even though the expression of the repressor tll in the right group is markedly going down over time, the level of hb does not exhibit a comparable change.
	Identify attribute similarities between populations	Also of interest is that while all the final levels of tll in the right group are the same,
	Identify attribute difference(s) between populations	the final levels of hb in the same set of cells varies. This latter observation is a possible hint that the final levels of tll do not matter for the expression of hb.
	Relate finding(s) to domain	Both of these observations point to potentially interesting features in how the hb gene is regulated, and are trends she will look to confirm in her statistical model.

3.15 NeuroLines Section 9.1

Goal	Step	Paper Text
Describe Observation (Item)	Isolate instance (input)	First, he explored the entire data set, sorted all neurites depending on neurite type and the number of synapses, to narrow down on a first structure of interest (i.e., dendrite D1). Using a visual query, the data set was reduced to only include dendrite D1, all its connected axons, and all dendrites these axons connect to.
	Identify attribute(s) to define/refine observation	Next, the scientist analyzed the detailed connectivity patterns, starting from dendrite D1. An initial analysis of the attributes of all synapses of this dendrite did not reveal any apparent patterns.
	Identify attribute(s) to define/refine observation	Therefore, the scientist first identified several multi-hit axons connected to dendrite D1 (Fig. 10), and then analyzed only the synapses between these axons and dendrite D1. Some of the attributes that the scientist looked at were given as scalar values (e.g., spine volume), while for other attributes (e.g., “closeness” of both neurites around the area of the synapse) the integrated 2D and 3D views of the original EM data were used. This allowed the scientist to further narrow down his analysis process and to slightly adjust and refine his hypothesis.
	Verify observation externally	When he was sufficiently sure of his findings he handed the data over to a statistician to conclude the analysis.

3.16 NeuroLines Section 9.2

Goal	Step	Paper Text
Describe Observation (Aggregate)	Identify attribute(s) to define/refine population(s)	To analyze synapse attributes in relation to the branching pattern of excitatory neurites, our collaborator started by identifying excitatory neurites.
Evaluate Hypothesis	Identify attribute difference(s) between populations	When the function of a neurite is unknown, analyzing the number of spine/shaft synapses gives an intuition for whether the neurite is excitatory or not. The scientist evaluated synapse properties close to the cell body in comparison to synapses far away from the cell body (but on the main trunk), and to synapses on far away branches.
	Assess hypothesis	In this particular case, the study was inconclusive
	Broaden population scope	and led him to acquire a bigger data set that will allow him to repeat this analysis with synapses spread out over a longer distance along a dendrite.

3.17 Paramorama Section 5.1

Goal	Step	Paper Text
Discover Observation	Note observation	The user began by using the step-through selection feature for the hierarchy shown in the overview: the next previously unseen subtree at level three (see Figure 6(a)) was accessed with a single mouse click to show a 4 × 4 matrix of outcomes (Figure 6(b)). By comparing these images and by moving the mouse over individual images to superimpose them on the reference image, the user easily identified poor results and tagged these (Figure 6(c))
Identify Main Cause (Item)	Examine finding(s) with other instances of observation	After identifying a few poor result sets
	Identify likely dominant cause	the user suspected that these poor outcomes occurred when parameter $p_4 = 13$, its highest sampled value.
	Assess hypothesis based on data change	The user was able to quickly confirm this by moving p_4 to the first level in the clustering hierarchy and by scanning all outcomes clustered under the last value for p_4 in the refinement view, none of which identified the two nuclei mentioned in the previous paragraph.
	Identify attribute(s) to define/refine population(s)	All outcomes for this value of p_4 were then tagged as negative, as shown in the bottom subtree of Figure 6(e). To reduce the complexity of subsequent analysis, the user next used the filtering facility of our prototype to hide all these negatively tagged outcomes.

3.18 Poemage Section 8.1

Goal	Step	Paper Text
None	Context	One collaborator described her approach to using Poemage in analyzing a poem as “noodling,” hovering over one sonic feature after another in the set view and poem view, selecting and deselecting rhyme sets almost arbitrarily. She said her greatest successes and insights came in every case when she happened on something indirectly, through idle play — as she says, “almost out of the corner of my eye.”
Discover Observation	Note observation	A specific example of this was an insight gained when glancing at the placement of nodes in the path view for the poem “Night” by Louise Bogan. While the placement of nodes in this poem is mostly regular in that there are generally a similar number per line (around 4) and they are mostly at similar distances from each other, indicating that there is typically about the same number of words per line and these words are of similar length, there was one line that had only two nodes, the second following very closely on the first. Thus, the abstracted view of poemspace revealed an immediately visible anomaly in the spatial distribution of words.
	Relate finding(s) to domain	This anomaly coincided with a powerful semantic moment in the poem, leading this collaborator to explore the rich sonic turbulence at that location and its connection and reinforcement of the semantic flow of the poem. She said that this view shed new light on a poem with which she was deeply familiar: “In other words, not only is this the poem’s turn, its pivot and crisis, but there’s just a whole lot going on, a lot I wouldn’t necessarily have considered in quite this way without the tool drawing my idle eye — a lot I hadn’t in all these years considered up to this moment.”

3.19 Ravel Section 5.1

Goal	Step	Paper Text
Discover Observation	Note observation	Fig. 10 shows the Vampir and Ravel visualizations of a complete 16 process, 4-ary merge tree. The logical step view is colored by lateness. In this case, the initial step is late, meaning the lateness is due to a severe load imbalance caused by the input data characteristics.
Collect Evidence	Form hypothesis	As process 0 is imbalanced, lateness is propagated to the corresponding gather as well as the resulting re-broadcast. Once the root of the tree is reached in the second gather stage lateness resets – there are no other events in the logical step with which to compare. The logical steps clearly highlight the gather tree structure of the algorithm and provide immediate insight into the overall behavior of the code. None of this information is easily

		accessible from the traditional trace visualization as the time-bound layout obscures the underlying structure.
	Identify evidence to support hypothesis	Figs. 11a and 11b show two more realistic cases – a 1,024 and 16,384 process run using an 8-ary gather with the same data. There is more lateness in the 1,024 run than the 16,384 process run. As in the small 16 process example, the lateness is due to load imbalance caused by the data. The 1,024 process run divides the data into larger portions per process than the 16,384 process run. The larger data portions are able to exhibit more variance in computational requirements, which is why the 1,024 process run experiences more lateness.
	Identify evidence to support hypothesis	Since 1,024 and 16,384 are not powers of eight, the level closest to the root only contains two gather processes at 1,024 nodes (and thus 16 processes at the next) and 4 gather processes at 16,384 nodes. In Fig. 11a, we can see a division in connection between the top and the bottom half of the image, with each half containing eight parallelograms, 16 groups in total. Similarly, the right side of Fig. 11b contains four large groupings.
Discover Observation	Note observation	However, the repeating motif of a parallelogram with a panhandle results in many timesteps where few processes are active. A closer analysis reveals a potential flaw in the algorithm. The panhandle occurs when a process sends to its leaves, those closest in rank space, before sending to its higher level children. Furthermore, this means the gather processes first send the information back to the leaves before sending it onward towards the root of the tree.
Identify Main Cause (Item)	Focus on instance	This motif manifests at each level of the tree, the largest example being the events in the middle of the logical steps on the higher half of the MPI ranks (lower half of the visualization).
	Identify likely dominant cause	The observed ordering misses an opportunity for a more aggressive pipelining of the computation. No process can finish until the root of the gather has been reached and thus the gather should be prioritized over the scatter
Collect Evidence	Form hypothesis	Upon discovering the non-optimal message ordering in Ravel, the merge tree developers changed their implementation. [Implicit hypothesis: problem was resolved with new implementation]
	Identify evidence to support hypothesis	Figs. 11c and 11d are traces of the improved application using the same input parameters, but showing significantly more overlap in the communication.
	Identify evidence to support hypothesis	For the large examples, the logical view can no longer meaningfully represent the messages, which are not drawn. The focus processes in the cluster view can be used to get a sense of the communication pattern. Fig. 12 features the Ravel cluster view for one phase of the improved merge tree algorithm run at both 1,024 and 16,384 processes. Though the two traces differ in scale, the messaging pattern conveyed is equally visible in both views.

	Identify evidence to support hypothesis	The focus processes show the optimized messaging pattern. The second process sends to three different groups of eight children, a group for each level of the gather tree for which that process is active. After the child receives the message, it in turns starts sending, except for the last group. This indicates the final group of eight which do not send represent the leaves, verifying the updated merge tree implementation handles the leaves last
	Identify evidence to support hypothesis	The coloring of the cluster visualization shows that the processes have been grouped by lateness in the aggregate non-communication steps late in the phase.

3.20 Ravel Section 5.2

Goal	Step	Paper Text
Collect Evidence	Form hypothesis	As such, we expected to see lateness vary among groups of 16 processes in the default mapping and then this lateness to propagate along the processes. We also expected the round-robin mapping would not have as much lateness variance in groups of 16 during the z exchange, but what small variance it did have would help alleviate contention further along the exchange where conflicts could occur, leading to lower lateness overall.
	Identify evidence to support hypothesis	In the 1,024 process scenario, Ravel shows behavior that supports this theory during the z exchange: the first four steps of Fig. 13a show a lot of variance in lateness between neighboring processes, visible in both the logical and clustered view, and lateness propagating along message lines, visible in the focus processes of the clustered view. The first four steps of Fig. 13d show no lateness as expected.
Discover Observation	Note observation	However, after the first four steps, we see unexpected gradient patterns along the processes under both mappings. These gradients are also present in the larger runs.
Describe Observation (Aggregate)	Identify attribute(s) to define/refine population(s)	The unexpected gradients are most visible in Figs. 13b, 13c and 13f. Also, rather than lateness propagating along a process, it inverts between the first two and the second two timesteps in these same examples.
Identify Main Cause (Item)	Examine finding(s) with other instances of observation	When we adjust the color map of the 8,192 process round-robin trace (Fig. 14), we also find similar problems.
	Assess finding(s) based on data	We verified that these observations were not drawing artifacts by examining focus processes in the cluster view and zooming in on the logical view. These effects could not be seen in a standard physical time visualization (Fig. 15).
	Identify likely dominant cause	In the first two send-receive pairs (timesteps 0 - 3), lateness starts at the send, but is not entirely caused by contention from the mapping, because that does not explain the gradient.

	Identify likely dominant cause	We hypothesize another contributor to this lateness. The MPI send call used can require the receiver to acknowledge the request. In the first two steps, only the last processes, those corresponding to the final xy plane in the domain decomposition, do not send. This is visible as the white space at the bottom of step 0 and the top of step 2. By not sending, those processes are able to more promptly respond to the request. Thus, their senders receive acknowledgement first and can complete their send. This effect cascades along the process IDs, resulting in the observed gradient. Fig. 16 illustrates this effect. The second send-receive timestep pair has the same problem inverted. In the 1,024 process runs, no process both sends and receives in the first two timesteps, so no cascade effect occurs which explains why no gradient appears.
	Examine finding(s) with other instances of observation	The send-receive timestep pairs in the y and x direction have similar boundary conditions, but are tiled in the process ID space. This leads to the striping patterns seen in some of the later steps. However, if the effect of the earlier delays is great enough, the effects of the later steps may not be significant enough to show up in this view.
Collect Evidence	Form hypothesis	The developers were previously unaware of this cascading dependency problem. They were able to fix the problem leading to a significant performance benefit. [Implicit: Fix will solve the problem]
	Identify evidence to support hypothesis	Fig. 17 shows the difference in time spent in communication for the full (non-benchmark) application before and after the change with all other parameters fixed.

3.21 SellTrend Section 7

Goal	Step	Paper Text
Discover Observation	Note observation	We can immediately identify in the Timeline View that a failure spike occurred on Monday, August 27th
Identify Main Cause (Aggregate)	Identify attribute(s) to define/refine population(s)	After selecting failed transactions over all 24 hours in that day...
	Identify likely dominant cause	We can see that, for example, travel agent DYS has contributed the most failed transactions in comparison to the other agents,
	Assess finding(s) with data	but the region representing it is a light shade of blue, indicating that it has contributed less failures in the selected time window than normal
	Identify likely dominant cause	Instead, we see that airline A80 is contributing a large volume of failed transactions,

	Assess finding(s) with data	worse than its historical average
	Identify attribute(s) to define/refine population(s)	We focus on airline A80 so only its failed transactions are shown.
	Identify likely dominant cause	Examining the other attributes, we note more than one to explore further, for example, class Z and R, flight number 4360 and error code 78 (Figure 7(c)).
	Identify attribute(s) to define/refine population(s)	From our previous experience, we decide to drill down to flight number 4360 first, because flight number is typically a key attribute.
	Identify likely dominant cause	It then becomes clear that airline A80 and flight number 4360 might be part of the cause of the failure spike on August 27th. Class Z and R together, and error code 78 correspond to 100% of failed transactions associated with A80 and Flight Number 4360.
	Assess finding(s) with data	This discovery is also corroborated by the overlaid information in the Timeline View, where we can see that August 27th stands out with a visible surge in failed transactions associated with these two attribute values
	Identify attribute(s) to define/refine population(s)	The hour panel shows that the surge was happening around 3-4pm and 9pm on that day.
	Identify likely dominant cause	The Attribute Map View further indicates that travel agent Z7F might be the right stakeholder to contact for resolution.

3.22 SignalLens Section 5.1

Goal	Step	Paper Text
Discover Observation	Examine attributes for unusual or interesting observations	We begin by observing that the overall signal appears to be quite uniform
	Note observation	except for a few instances where a slight “whisker” appears to protrude from the otherwise uniform shape. Figure 4 shows an example of such a whisker and the sequence of navigation events that narrow into the corresponding signal glitch. Figure 3A shows the lens positioned over this feature.
Describe Observation (Item)	Examine finding(s) with other instances of observation	We can use the signal motif finder to look for a similar wave shape as this glitch in Figure 3A. After filtering for only statistically significant matches we plot the top track show in Figure 3C
	Identify attribute(s) to	At this stage we notice that the repetition of the glitch generally appears every other carrier oscillation except for a few instances

	define/refine observation	which seem to skip this pattern.
	Examine finding(s) with other instances of observation	We can investigate further by examining 1st and 2nd derivatives. Filtering for just the most extreme slopes and curvatures yields the additional two tracks shown in Figure 3C.
	Identify attribute(s) to define/refine observation	In this case we notice that the pattern is uniform and inspection of the indicated locations show that even when the motif finder missed a glitch, the filtered derivatives indicate its location

3.23 SignalLens Section 5.2

Goal	Step	Paper Text
Discover Observation	Calculate derived attributes	We begin by calculating rise and fall times of the signal pulses. These calculations provide us the location of the leading and trailing edges of the PCI-E pulses. We calculate a third track from the first by calculating the distance from the rise to fall time, essentially giving us the pulse width.
	Examine attributes for unusual or interesting observations	If we then view the histogram of pulse widths (Figure 5E) we notice they mostly follow the expected equally spaced distribution of PCI-E widths corresponding to the different width multiples in the specification.
	Note observation	However, we notice that there is a very narrow outlier peak.
	Examine finding(s) with other instances of observation	Using the histogram, we can filter for just this peak. Navigating to the single outlier, we find an extremely sharp spike at 2.1426 ms as shown in Figure 5A.

3.24 SnapShot Section 5.1

Goal	Step	Paper Text
Discover Observation	Note observation	Our first analyst (A) had been investigating shot lengths at a specific hockey arena, Madison Square Garden in New York. He had been doing so by studying raw data and game footage. He began his visual exploration with SnapShot by creating small multiples of each rink in the league (Figure 1) then increased the band-width to 10-feet (Figure 16). He remarked, "I will be able to see right away if this rink really is different than others." As evidenced in the figure, he found that NYR (Madison Square

		Garden) does show a different pattern of shots than any of the other rinks.
Describe Observation (Aggregate)	Identify attribute(s) to define/refine population(s)	He SnapShotted six traditional heat maps: home shots only and away shots only for each of the three rinks. He told us that he was only concerned with even strength shots and filtered out power play shots and short handed shots.
Identify Main Cause (Item)	Identify likely dominant cause	After studying the relative differences in shot patterns for the three teams, he changed the views to shot maps so that he could color shots by player position. He was interested in identifying which positions might be contributing to the atypical shot patterns.

3.25 SnapShot Section 5.2

Goal	Step	Paper Text
Evaluate Hypothesis	Form hypothesis	Analyst B began the session by declaring, “Animals defend their homes”. He was interested in observing if hockey teams follow this ‘rule of nature’: do they play more defensively at home and offensively on the road? He hypothesized that “playing defensively” would imply longer even-strength shooting, because “the home team is not willing to give up any ground by over-attacking on offense”.
	Identify attribute(s) to define/refine population(s)	As his hypothesis was a high-level generalization, he chose to keep each filter at the highest level, and never drilled down to any specific teams, rinks, or times. Furthermore, because shot length was far more important to him than specific locations, he chose to explore exclusively using radial heat maps. He SnapShotted two radial heat maps (Figure 17, top row): home shots and goals and away shots and goals, each at a two foot granularity. Because the views included all thirty teams, he mentioned that there may be several teams skewing the distribution of shot lengths, but that he still expected to find slight differences in shooting patterns.
	Identify attribute difference(s) between populations	By toggling back and forth between the two radial heat maps, the analyst noticed and pointed out two rings on the ice that interested him. While both heat maps had a red ring around the thirty foot shooting mark, the ring is comparatively darker on the visitor team map (each ring indicated with a arrow arrow in the top row of Figure 17).
	Identify attribute difference(s) between populations	Additionally, the dark ring fifty feet from the net on the home map (green arrow in Figure 17) was not as pronounced on the away map.
	Relate finding(s) to domain	These two insights caused the analyst to theorize out loud, “More long shots from home teams and less of those thirty footers, I wonder which of these are actually goals?”
	Identify attribute	This query led him to SnapShot two additional radial heat maps (Figure 17, bottom row): one of home goals and one of away goals.

	difference(s) between populations	
	Identify attribute difference(s) between populations	The distributions appeared to be nearly identical.
	Assess hypothesis	This excited the analyst. He concluded that “They are shooting differently at home but they aren’t scoring differently!” As he ended his evaluation, the analyst asked if he could access the system in a non evaluation capacity and use these visualizations in a client presentation about the “Animals defend their home” theory.

3.26 SnapShot Section 5.3

Goal	Step	Paper Text
Evaluate Hypothesis	Form hypothesis	The third evaluator, Analyst C, chose to use SnapShot to investigate potential “sweet spot” areas on the ice, those areas with a high conversion rate of shots to goals. He commented, “Since most goalies have their stick in their right hand and glove in their left, I am wondering if they struggle to move left to right, and if shooters are capitalizing on this.” Essentially, he was hypothesizing that more goals result from shots on the goalie’s right side than left side.
	Identify attribute(s) to define/refine population(s)	To explore this possible phenomenon, the analyst created a shot map colored by shots and goals. He quickly turned off the blue (failed) shots and displayed only the red goals (Figure 18, top left).
	Identify attribute differences between populations; Assess hypothesis	His initial impression was that there was little evidence that goals came from non-symmetrical locations on the ice
Evaluate Hypothesis	Identify attribute(s) to define/refine population(s)	so he created several shot maps based on his domain knowledge of hockey dynamics. One displayed goals by just left and right wingers (Figure 18, top right); one displayed goals during the third period (Figure 18, bottom left); and one displayed only goals scored from January onward (Figure 18, bottom right).
	Form hypothesis	As he explored each image, he talked about theories common amongst NHL coaches such as “intuition says that a shot has a bigger advantage if the goalie isn’t set yet”.
	Identify attribute differences	None of the four shot maps that he created (and the corresponding traditional heat maps he created later) supported his hypothesis that players scored more goals shooting from the goalie’s right

	between populations; Assess hypothesis	side.
None	Context	However, this dead end only intrigued the analyst more. In his words, "I like to feel that I'm being pushed to think more creatively. This was my first opportunity to explore these theories." He mentioned that he would like to continue to search for "soft spots" using the system in the future, particularly areas outside of "the house", the term he used to describe the high-trafficked pentagon defined by the net, faceoff circles, and the blue line.

3.27 SoccerStories Section 6.1

Goal	Step	Paper Text
Discover Observation	Note observation	"The Offensive Defender": He began the first article during his initial exploration of all the phases when he was surprised to see that Real Madrid's defender Varanne (number 2) was, despite his nominal role, active in many offensive phases of the first game.
Describe Observation (Aggregate)	Identify attribute(s) to define/refine population(s)	To illustrate this, he selected Varanne to highlight his actions and took the screenshot shown in Figure 14(a).
Explain Differences	Identify attribute difference(s) between populations	Proceeding with his analysis, he found out that this player was much less involved in offensive phases in the second game.
	Identify attribute difference(s) between populations	He also compared Varanne's statistics in both games, which showed that the player made much more passes (48) in the second game than in the first (33).
	Relate finding(s) to domain	Based on what he found out with SoccerStories and his previous knowledge, he deduced that Varanne (and to some extent the whole Real Madrid team) performed this way due to the location of the games: when not playing at home they preferred to wait for the other team to make a risky move and then counter-attack.
Discover Observation	Note observation	"Rewarded Coaching": The analyst began working on his second article when he inspected the small multiples for the second game and saw that phases during its last third contained significantly more actions.
Evaluate Hypothesis	Form hypothesis	This immediately reminded him that during this game Manchester United's player Nani received a red card around the 60th minute and that Real Madrid's coach opportunistically substituted defensive player Arbeloa for the offensive Modric. [Implicit hypothesis: the change in the number of actions was caused by a player substitution -- Modric substituted in -- (which he

		remembered had occurred in this game)]
	Identify attribute(s) to define/refine population(s)	He explored the phases following the substitution
	Identify attribute difference(s) between populations	and noticed that Modric became very present in both phases (6 of a total of 11 remaining phases) and actions (an average of 3.5 actions per phase).
	Identify evidence to support hypothesis	Modric also scored the goal that tied the game and was key to his team's second goal, which guaranteed its victory.

3.28 TenniVis Section 5.2

Goal	Step	Paper Text
Discover Observation	Note observation	When analyzing the women's tennis match, the coach only spent a few minutes on the Pie Meter view (which is the default view when the application is started). She noticed the very high number of service breaks as indicated by the red and green boxes. She then switched to the Fish Grid view (see Figure 7).
Describe Observation (Aggregate)	Identify attribute(s) to define/refine population(s)	After scanning through all of the points in order in the Fish Grids, she stopped to more closely examine games three and four in the second set (Figure 7(a) and (b)).
Identify Main Cause (Item)	Focus on instance	When zoomed into the second semantic zoom level, she was able to see the outcomes of each point in these games (represented by lowercase letters).
	Identify likely dominant cause	In game three, she noticed from the Fish Grid how her player (player one) broke her opponent's serve with three solid shots (i.e., two forced errors and one winner) plus an unforced error made by her opponent. She then noted, however, that her player committed two double faults and two unforced errors in her next service game (game four).
Collect Evidence	Form hypothesis	Based on prior experience with her player, the participant suspected that her player may have committed the two unforced errors by trying to hit too many down-the-line shots (a potentially risky shot) versus going cross-court. She commented that "I've really been trying to work on her [the player] hitting the ball cross-court more because she goes down the line and then she either misses it or then they make her run cross-court. . . so I was immediately thinking 'How did she lose this point?'".
	Identify evidence to support hypothesis	She loaded several points one-by-one into the video player and was able to find several examples that confirmed her hypothesis.
	Identify	She then decided to focus on serving and, using the zoom slider, further zoomed into game four to see more details such as first

	attribute(s) to define/refine population(s)	serves vs. second serves.
	Identify evidence to support hypothesis	From Figure 7(b), she noticed that five of the six points were from second serves (including two double-faults).

3.29 TenniVis Section 5.3

Goal	Step	Paper Text
Describe Observation (Aggregate)	Identify attribute(s) to define/refine population(s)	Once the data set for his own player was brought up,
Explain Differences	Identify attribute difference(s) between populations	he immediately looked at the histogram (see Figure 8 (a)) coupled with the Pie Meter view to get an overall sense of the difference between the two players in terms of point outcomes. He commented that “overall, their guy hit certainly more winners. . . he also had more forced errors and our guy had less unforced errors. . . our guy doubled less and had a few more aces. . . but that looks a little negligible in terms of risk/reward”.
	Identify attribute difference(s) between populations	He then began using various filter combinations to create individual histogram snapshots for comparison, noticing that his player had a lot of points that started from his second serve.
Discover Observation	Note observation	Continuing his analysis after switching to the Fish Grid view, he noticed a key game in set two when his player was serving at five games to three (see Figure 8 (b)).
Identify Main Cause (Item)	Focus on instance	He focused in on the point when his player was up 30-15 (only two points away from winning the set) and his opponent hit a winner (see Figure 8 (b)). The opponent then was able to regain the upper hand in the match and ultimately win it.
	Identify likely dominant cause	He analyzed why his player lost this point by selecting it to view in the video player. He noticed that, although his player served his opponent with a tough serve out wide, he failed to capitalize on this advantage and gave his opponent an easy putaway shot at the net.
	Relate finding(s) to domain	This insight led the coach to indicate he would discuss shot selection with his player to avoid giving away the momentum in a match.

3.30 Vials Section 6.1

Goal	Step	Paper Text
Describe Observation (Aggregate)	Identify attribute(s) to define/refine population(s)	In this case study, the experts chose to compare variants of the SRSF7 gene using samples of a brain cancer (glioblastoma, GBM, 100 samples) and a form of leukemia (acute myeloid leukemia, LAML, 167 samples), which corresponds to Goal G1—exploring differences between samples and groups. The gene SRSF7 regulates alternative splicing at a variety of targets genome-wide, while the gene itself is also regulated by alternative splicing. They chose to investigate SRSF7 because exon 4 shows large differences in how often it is “used” in LAML and GBM. Because differences in usage are derived from measurements of both exon expression levels and data about junctions across the two diseases, this gene is a good analysis target for the described tasks (see Section 3).
Evaluate Hypothesis	Form hypothesis	[Implicit: both exon abundance and junction use support a difference in exon 4 splicing between LAML and GBM]
	Identify attribute(s) to define/refine population(s)	Figure 13 shows SRSF7 with GBM samples highlighted in orange and LAML samples highlighted in blue. The TCGA data provides expression data as an average for every exon, as is evident from the constant blocks in the expression view in Figure 13. The expression data in Figure 13 is aggregated into the two disease groups LAML and GBM. When exploring this data, the domain expert noted that there is roughly equivalent expression of the exons that are not alternatively spliced between the two groups (e.g., exons 3,5,6,7).
	Identify attribute differences between populations	In contrast, the alternatively spliced exon 4 shows very low expression in GBM, but some expression in LAML (task T3 applied to the expression data).
	Identify attribute differences between populations	Consistent with this, our collaborators observed in the junction view (top), that there is greater support for the junction joining exon 3 to exon 4 in LAML (blue) than GBM (orange) (Fig. 13, pattern p1; task T3 applied to junction support).
	Identify attribute differences between populations	On the other hand, GBM samples show more support for the splice junction that skips exon 4 (higher orange values in pattern p2).
	Assess hypothesis	This confirms that both exon abundance and junction use support a difference in exon 4 splicing between LAML and GBM.
	Identify evidence to support hypothesis	Additional evidence is visible in the isoform abundance view for the first isoform (pattern p3). This isoform is characterized by the inclusion of exon 4 (task T3 applied to isoform abundance,

		combined with tasks C1 and C2). As expected from exon and junction data, this isoform is more abundant in LAML samples than in GBM (the blue dots show larger values than the orange dots in p3)
Describe Observation (Aggregate)	Identify attribute(s) to define/refine population(s)	<p>In addition to these differences, exon 8 of SRSF7 in the TCGA data is known to have weak but statistically significant alternative splicing [24]. While the support for differences in expression of this exon between GBM and LAML is small, exploring the junctions associated with exon 8 lead our collaborators to a new hypothesis regarding a yet unknown exon variant.</p> <p>Specifically, in ranking samples by support for the exon 8 – exon 9 junction (p4 in Figure 12), they observed that levels of non-alternatively-spliced junctions are generally highly correlated with this junction in both cancer types (both orange and blue samples show approximately equivalent correlation in the boxed scatterplots in Figure 12(b)).</p> <p>In contrast, two visually distinct populations emerge in the exon 2 – exon 3 junction (p5).</p>
Explain Differences	Identify attribute differences between populations	Specifically, LAML samples (blue) show a greater and apparently linear relationship with the exon 8 – exon 9 junction,
	Identify attribute differences between populations	while GBM samples (orange) display lower exon 2 – exon 3 junction use proportional to the exon 8 – exon 9 junction in the starred scatterplot.
	Relate finding(s) to domain	One potential explanation for this observation is an alternative transcription start site which is absent from the gene reference database, and which does not use exons 1 or 2, leading to an observation of the type of Goal G2—discover novel isoforms. This hypothetical alternative starting exon would be connected with a junction to exon 3

3.31 Vials Section 6.2

Goal	Step	Paper Text
Discover Observation	Note observation	While exploring the data, our collaborators were continuously looking for issues of data quality (goal G4). They eventually found a striking case of missing and wrong data in the gene EGFR in the Bodymap4 dataset. Figure 14 shows a case where the white blood cell sample, highlighted in red in Figure 14(a), shows strongly deviating behavior from other samples in the isoform abundance view.

Collect Evidence	Form hypothesis	Initially intrigued, our collaborators quickly identified that this is a data quality issue,
	Identify evidence to support analysis	as there is no expression data available for the white blood cell sample, as is evident when inspecting the data in the expression view, shown in Figure 14(b).
	Identify evidence to support analysis	Similarly, there is no junction support for this sample, indicating that the reported isoform abundances are an artifact of the processing pipeline. While this is an extreme case of a data quality problem, it would not be immediately apparent when only the isoform abundances are investigated.

3.32 Weaver Section 8.1

Goal	Step	Paper Text
Explain Differences	Overview data	He stated that he needed a sense of the spread or variation across the ensemble,
	Find attribute difference(s) between populations	but more importantly, he also needed to be able to understand how that variation differs from a particular model or member. He explained that organizations such as the National Weather Service still key their recommendations off a deterministic forecast, so understanding how the rest of the ensemble compares to that particular member is incredibly important. As such, he appreciated being able to interactively highlight a particular member from the ensemble. He was also particularly impressed by the contour boxplot summarizations. He stated that, while it would take training for forecasters to understand exactly what they are looking at, the contour boxplots provide the same visual cues of the forecast as spaghetti plots, but much more quickly and concisely.
Collect Evidence	Form hypothesis	[Implicit: visualization showed correct data]
	Identify evidence to support hypothesis	Using these views he was able to determine that the forecasts showed the expected signals for a threat of a lightning-started fire: moisture, indicating lightning potential, on the front end of the forecasts, followed by windy, dry, unstable conditions for a day or so after.
	Identify evidence to support hypothesis	Finally, our collaborator looked at the probability view in order to determine which areas had a high likelihood of a critical combination of dry, windy, and unstable conditions in the latter portions of the forecast. As we have reproduced in Figure 6, the combined condition of surface temperatures greater than 60° F, surface wind speeds greater than 20 mph, and a Haines index of 5 or greater highlighted the area over northern New Mexico as favorable for fire spread after lighting ignition. Our collaborator

		noted that this highlighted area, which he would have been worried about, is where the Diego fire originated.
--	--	---

5 Excerpt from the video analysis in which we applied our framework

Event transcribed from the video	Munzner [2014]	Schultz <i>et al.</i> [2014]	Our own words	Analysis Goal Framework
Loads dataset of census data of people who attend burning man.				
Says she knows it is survey data so you probably want to see something vs. something. Creates chart of gender vs. # of previous burns.	Analyze>Consume >Discover Search>Browse Query>Compare>At tributes>Many(2)> Correlation	compare variable distributions (exploratory, compare, distributions, attri(*) attri(*), all)	what distinguishes category X?	Explain Differences: She is trying to explain the difference between people who attend a lot and those who attend infrequently. Looking for the factors that differ between these populations.
Notices more male repeat customers than female repeats. But also speculates that there are more male attendees. Produces a chart to relate this idea to number of attendees overall.	Analyze>Consume >Discover Search>Lookup Query>Compare>At tributes>Many (2)?	compare variable distributions (exploratory, compare, distributions, attri(*) attri(*), all)	explore relationships	
Asks what is the difference between people who go to burning man a lot vs. those who do not go regularly. Copies previous worksheet, then swaps gender with religion.	Analyze>Consume >Discover Search>Browse Query>Compare>At tribute>Many (2)>Correlation	compare variable distributions (exploratory, compare, distributions, attri(*) attri(*), all)	explore relationships	
<break, then revisits previous chart> Notices that most attendees have no religion, but there are a significant number of Buddhists.	Analyze>Consume >Discover Search>Lookup Query>Identify>Attr ibute>One>Distrib ution	compare variable distributions (exploratory, compare, distributions, attri(*) attri(*), all)	distribution, high frequencies	

<break, discussion of other topics>				
Loads new dataset with her own personal music listening history. Doesn't know yet what she wants to see in it.				
Asks - How much did I listen to music, week by week? Plotted timeline for the year. Noted an anomaly (low point) and realized this was when she attended a conference and wasn't at work listening to music. The next week was a spike when she did nothing but listen.	Analyze>Consume >Discover Search>Browse Query>Identify>All Data>Trends, Features	Analyze trends (exploratory, *, trends, attri(*) attri(*time), all)	temporal trend	Describe Observation (Aggregate): She noted that sometimes she listens more than other times. Then aimed to fully characterize when she listens more.
Asks - What about day by day? Made a chart but it was too messy so she discarded it.	Analyze>Consume >Discover Search>Browse Query>Identify>All Data>Trends, Features	Analyze trends (exploratory, *, trends, attri(*) attri(*time), all)	temporal trend	
Asks - Which day of the week do I listen the most? Noticed it was Tuesday and explained that it was because she has few meetings that day. Friday was low – she doesn't stay in the office as long.	Analyze>Consume >Discover Search>Locate Query>Identify>Attr ibute>One>Distribut ion>Extreme	analyze trends, periodicity, (exploratory, *, trends/frequencies, attri(*) attri(*time), all)	temporal trend	
Asks - What are the times when I listen the most on different days of the week? Most listened time was Tues afternoon. Observed that she never came to work earlier than 8am.	Analyze>Consume >Discover Search>Locate Query>Identify>Attr ibute>One>Distribut ion>Extreme	analyze trends, periodicity (exploratory, *, trends, attri(*) attri(*time), all)	temporal trend, opportunistic observation	
<progressed to a new set of tasks related to what music she listened to>				

6 Excerpt from an analysis log file in which we applied our framework

(Cells with yellow backgrounds are not in the original framework but were added in this analysis)

Coders' Annotations	Goal	Step
Investigate the proportion of people executed by race over time (i.e., number of records / race / total people executed per time period).	Discover Observation	Examine attributes for unusual or interesting observations
Renamed the worksheet to "More white people than black are executed."		Note observation
Explore the distribution of Method of execution over time. Instead of by proportion, looked at counts. Also emphasize each individual (Name in detail shelf).	Discover Observation	Examine attributes for unusual or interesting observations
Interesting that the sheet is named "Most recent FIRING SQUAD". Sounds like he was trying to find an interesting angle of the data since lethal injection is expected but firing squad is more dramatic.		Note observation
Explored the age of the executed. Started with Juvenile (Yes/No), added a bunch of details (Age, Number / Race / Sex) to get more info in the tooltips to enable exploration.	Compare Entities	Identify attribute difference(s) between populations
Decided to look at Sex more explicitly but probably not enough females to make a meaningful comparison. Undo to get back the Juvenile/Name state.	Compare Entities	Examine if populations are appropriate for comparison
Further explored Age distribution over time	Discover Observation	Examine attributes for unusual or interesting observations
Formatting to make the chart resemble the Iraq's Bloody Tool chart	N/A: FORMATTING	
Explore the number of executed at a different level of detail (YEAR -> MONTH). Doesn't look like the shape of the bar distribution changed too much, so back to the cleaner YEAR.	Discover Observation	Examine attributes for unusual or interesting observations
Explore a bunch of bucketing to see how the time trends compare between the buckets. Foreign National field (Yes/No) and Federal didn't have enough Nos for a trend to form.	Compare Entities	Examine if populations are appropriate for comparison
"Region" and "Method" seemed promising but not pursued further, probably because of the lack of resolution (one population dominated).	Compare Entities	Examine if populations are appropriate for comparison
Finally focused on "State"	Compare Entities	Identify attribute difference(s) between populations

and found that Texas was an outlier. Renamed the sheet to "Way to go Texas".	Discover Observation	Note observation
Review the vizzes created. May be trying to see which would go into the dashboard / if he needed more vizzes for the dashboard?	N/A: DASHBOARDING	
Further explore time trends at the Monthly level. May be to find trend here (by Month); may be to find a good way to display the data at a lower level of detail.	Discover Observation	Examine attributes for unusual or interesting observations
In any case, back at the YEAR level to explore Race and State	Discover Observation	Examine attributes for unusual or interesting observations
and settled on State and explored the states that did NOT have a lot of executes (lower bound). Renamed the sheet to Rare Executors. Added details to show who the executed were.	Discover Observation	Note observation
Explore ways to display the barchart	N/A: FORMATTING	
Explore the number of records by method	Discover Observation	Examine attributes for unusual or interesting observations
and formatted the display.	N/A: FORMATTING	
Explore the distribution of race	Discover Observation	Examine attributes for unusual or interesting observations
Trying to add another dimension to the barchart. I think he is mimicking the Iraq's bloody toll visualization where there was # killed were showed as "coalition deaths" and "civilian deaths". He tried to add gender information but somehow decided against it.	Compare Entities	Examine if populations are appropriate for comparison
Explored Race distribution overall.	Discover Observation	Examine attributes for unusual or interesting observations
Explore Age distribution overall.	Discover Observation	Examine attributes for unusual or interesting observations
Experimented with ways to show the info.	N/A: FORMATTING	
Back to formatting the Race distribution barchart.	N/A: FORMATTING	
Formatting the total death chart.	N/A: FORMATTING	
Explored individuals. Maybe wanted to make sure the records are indeed unique (i.e., no duplicates to mess up his aggregates).	Collect Evidence	Identify evidence to support hypothesis