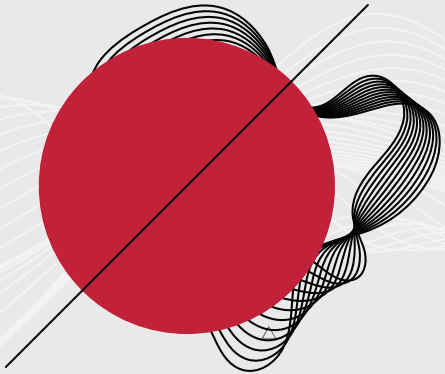
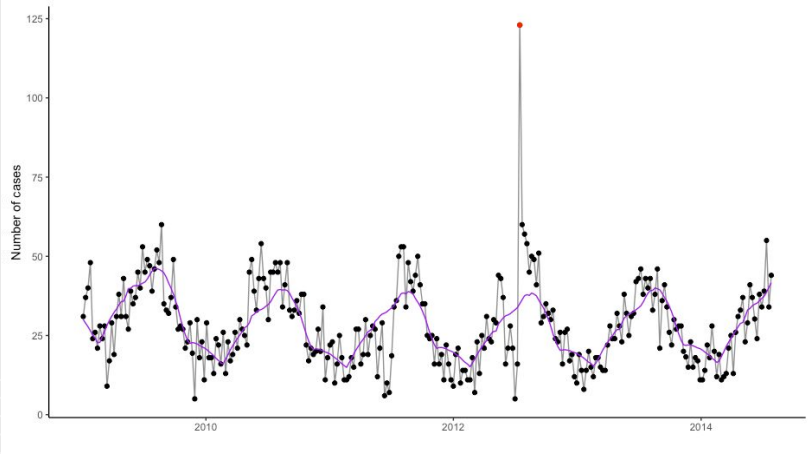
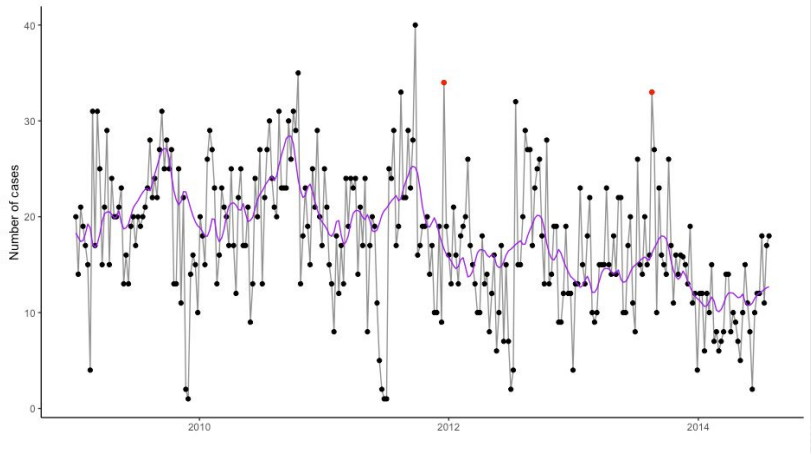
Disease Outbreak Radar:
A Tool for Public Health Users

UBC CPSC 547 Project
Cloris Feng, Derek Tam, Tae Yoon (Harry) Lee
Dec 10, 2020



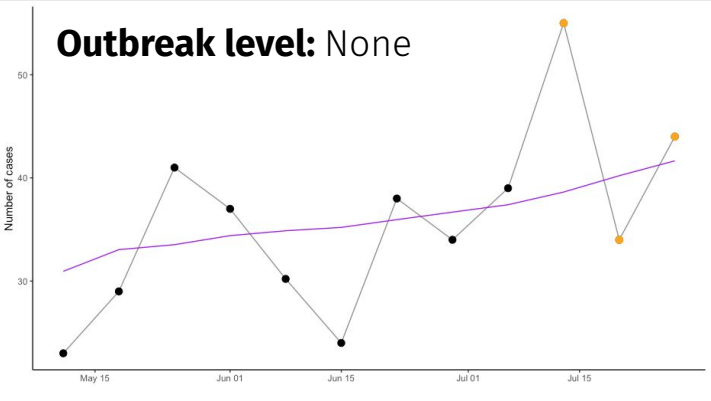
Challenges in Disease Outbreak Detection

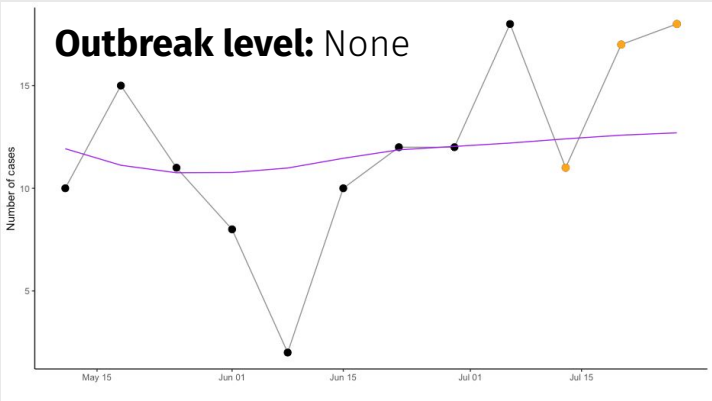




- “A disease outbreak is defined as the occurrence of disease cases in excess of normal expectancy.” [World Health Organization]
- Due to complex disease characteristics, it is difficult to define the “norm.”
- An automated method has been developed in collaboration with British Columbia Centre for Disease Control to reduce the burden and to help them identify disease outbreaks.

Outbreak level





- Outbreak p-value: the probability of observing the **three** most recent numbers of cases.
- Outbreak level:
  - High: outbreak p-value <= 0.1%
  - Medium: 0.1% < outbreak p-value <= 1%
  - Low: 1% < outbreak p-value <= 5%
  - None: outbreak p-value > 5%

Motivation

BC CDC is monitoring around **60** diseases in each of the **16** regions of BC. They need to check whether there is an outbreak in 60*16 ~= **800** diseases **weekly**.

WHAT: Multidimensional Table

Publicly available disease count data of the United States was used (Project Tycho).

Attribute name	Attribute type	Levels/range	Description
Disease	Categorical	6	Disease name
Week date	Sequential/ ordered	2009-01-04 - 2014-07-27	Start and end dates of a week in which the number of cases is collected (~290 observations)
Region	Categorical	9	US census divisions
Number of cases	Quantitative	0 - 3,140 (integers)	Number of cases of a disease
Number of population	Quantitative	14,469,650 - 62,382,273	Number of population estimated by the US Census in the year of Week date
Derived attributes			
Rate of cases	Quantitative	0.16 - 1094.90 (per 10,000,000)	Number of cases divided by the number of population
Estimated number of cases	Quantitative	Non-negative real valued	Estimated by the method
Outbreak p-value	Quantitative	0-1 (real-valued)	Probability of observing the 3 most recent numbers of cases based on the method
Outbreak level	Categorical	User-defined	Level of outbreak specified by public health users based on the outbreak p-value

Table 1. List of attributes and derived attributes.

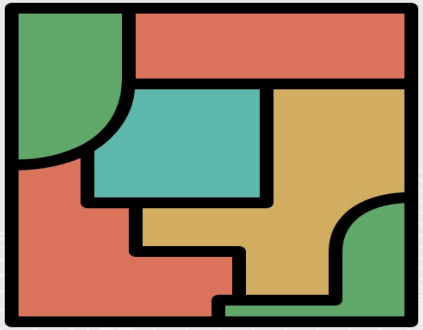
WHY: Search, Analyze, and Compare

Task 1

Efficiently **search** for an outbreak in diseases across regions

Task 2

Analyze the trend in the number of cases and estimated number of cases for each disease with an outbreak

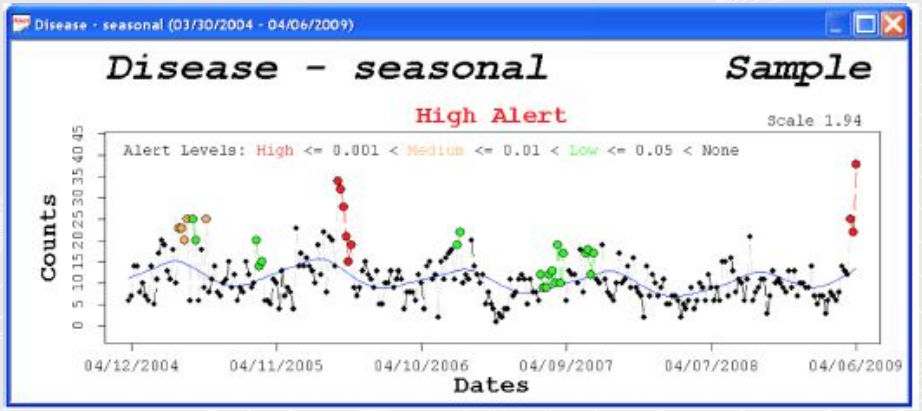


Task 3

Compare the outbreak levels across regions for each disease with an outbreak

Analysis of COVID Dashboards and Design Guidelines

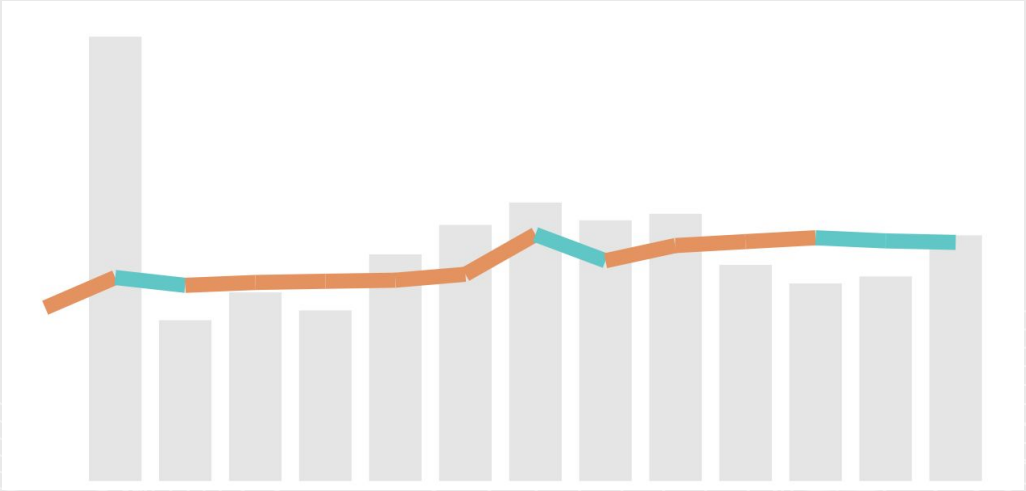
- Analyzed 4 COVID dashboards & 8 design guidelines
  - Learned lessons specific to visualization of **case data** with **geographical context** and applied them to our prototype
  - Guidelines were used to specifically address Task 2 and Task 3



Recommendation for Task 2

- **Task 2**:
  Analyze trends in the number of cases for a given region and disease

- **Recommendation**:
  A line-over-bar chart, with gain/loss trend-lines and annotated weeks of interest

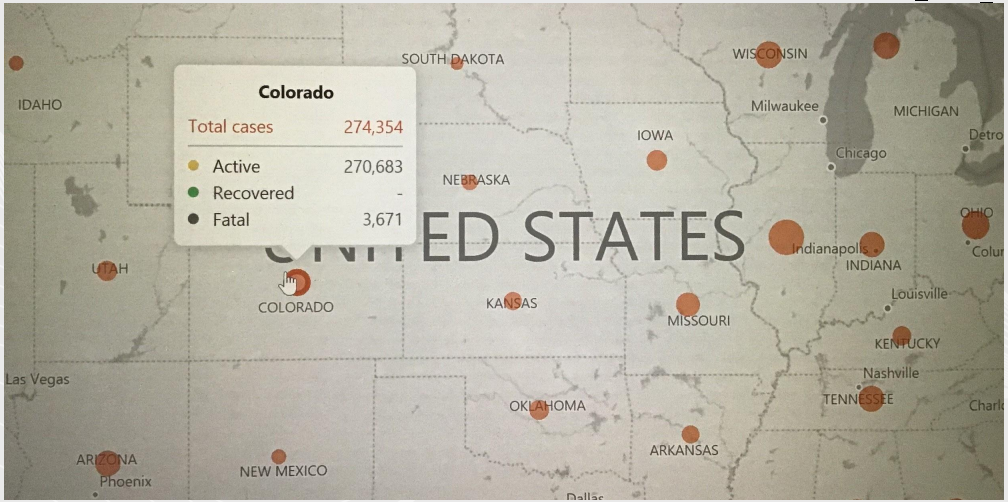


"Track Testing Trends," *Johns Hopkins Coronavirus Resource Center*. https://coronavirus.jhu.edu/testing/tracker (accessed Nov. 17, 2020)

Recommendation for Task 3

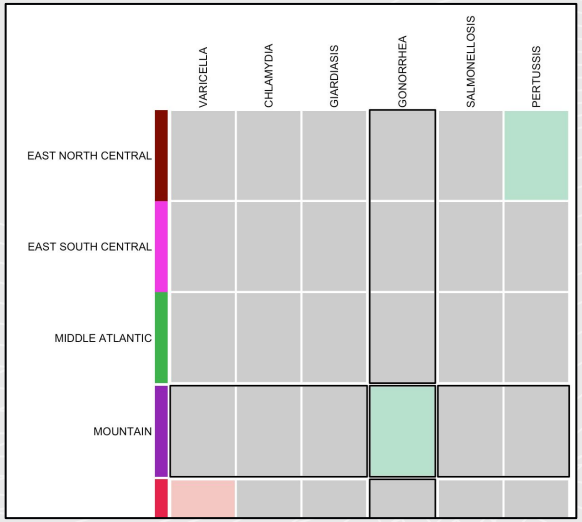
- **Task 3**:
  Compare the outbreak risk across regions for each disease

- **Recommendation**:
  A combination of bubble map and dot-density map to handle case number complexity



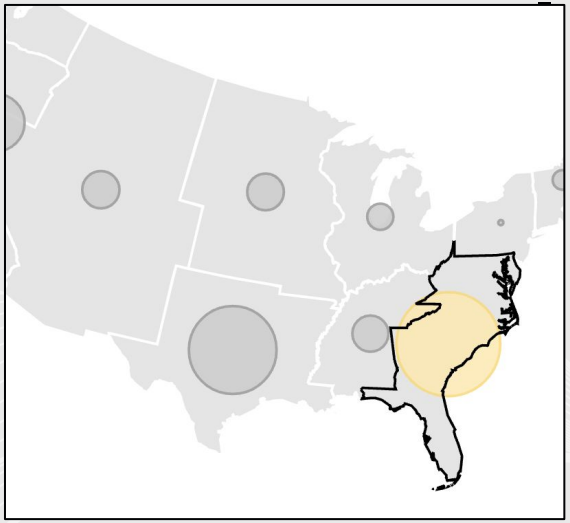
"Microsoft Bing COVID-19 Tracker."
http://bing.com/covid/local/maine_unitedstates?dynamicSharing=true&shtp=Facebook&shwth=900&shh=800&shtk=Y29yb25hdmlydXMgdHJhY2tlciB1cGRhdGVz&shdk=dGVzdA%3D%3D&shth=OSH.Mmq%2BwuM5WWI/TcdViNGxBA&redirect_uri=http%3A//veeraux%3A81/covid/local/unitedstates%3FdynamicSharing%3Dtrue&ref=Coronavirus&al (accessed Oct. 22, 2020).

HOW: Implementation Design



Task 1

Task 2



Task 3

Implementation and Demo

- Outbreak detection algorithm was developed by Harry as part of his Master's thesis
  - Algorithm was applied to a subset of Tycho data

- Dashboard was implemented in **R**
  - Made interactive with the **ShinyApp package** and **plotly**
  - Web-hosting through **shinyapps.io**

Limitations and Next Steps

- Tycho dataset is an **adequate but non-ideal** proxy for real data
- Dashboard could use additional quality-of-life features
  - Reordering and removal of individual timeseries
  - Region selection via map
- Reporting and export for diseases and regions of interest