

LendingClub CT analysis

Andrew Ba Tran

May 16, 2016

This is the methodology used for the story: What do Connecticut residents borrow for from online lenders?
(<http://trendct.org/2016/05/17/what-people-took-out-loans-for-in-connecticut>)

Visit the repo (<https://github.com/trendct/data/tree/master/2016/05/stanford-cepa/lending-club>) for the data used in this analysis or visit Kaggle's data site (<https://www.kaggle.com/datasets>) for the Lending Club data (<https://www.kaggle.com/wendykan/lending-club-loan-data>) and other scripts that analysts have submitted (<https://www.kaggle.com/wendykan/lending-club-loan-data/scripts>).

```
library(dplyr)
library(tidyr)
library(choroplethr)
library(stringr)
# devtools::install_github("hrbrmstr/ggalt")
library(ggplot2)
library(ggalt)
library(scales)
```

Mapping national data

```
loans <- read.csv("data/loan.csv", stringsAsFactors=FALSE)

us_map <- loans %>%
  group_by(addr_state) %>%
  dplyr::summarise(loans=n(), total=sum(loan_amnt), average=round(mean(loan_amnt),2))
```

```
## Warning in summarise_impl(.data, dots, environment(), caller_env()):
## integer overflow - use sum(as.numeric())

## Warning in summarise_impl(.data, dots, environment(), caller_env()):
## integer overflow - use sum(as.numeric())

## Warning in summarise_impl(.data, dots, environment(), caller_env()):
## integer overflow - use sum(as.numeric())

## Warning in summarise_impl(.data, dots, environment(), caller_env()):
## integer overflow - use sum(as.numeric())
```

```
# Bringing in US population
uspop <- read.csv("data/uspop.csv", stringsAsFactors=FALSE)

us_map <- left_join(us_map, uspop)
```

```
## Joining, by = "addr_state"
```

```
# Adjusting per capita
```

```
us_map$loans_per_capita <- round(us_map$loans/us_map$population*10000,2)
```

```
us_map <- us_map[c("state", "loans_per_capita", "loans", "total", "average")]
```

```
us_map$state <- str_to_lower(us_map$state)
```

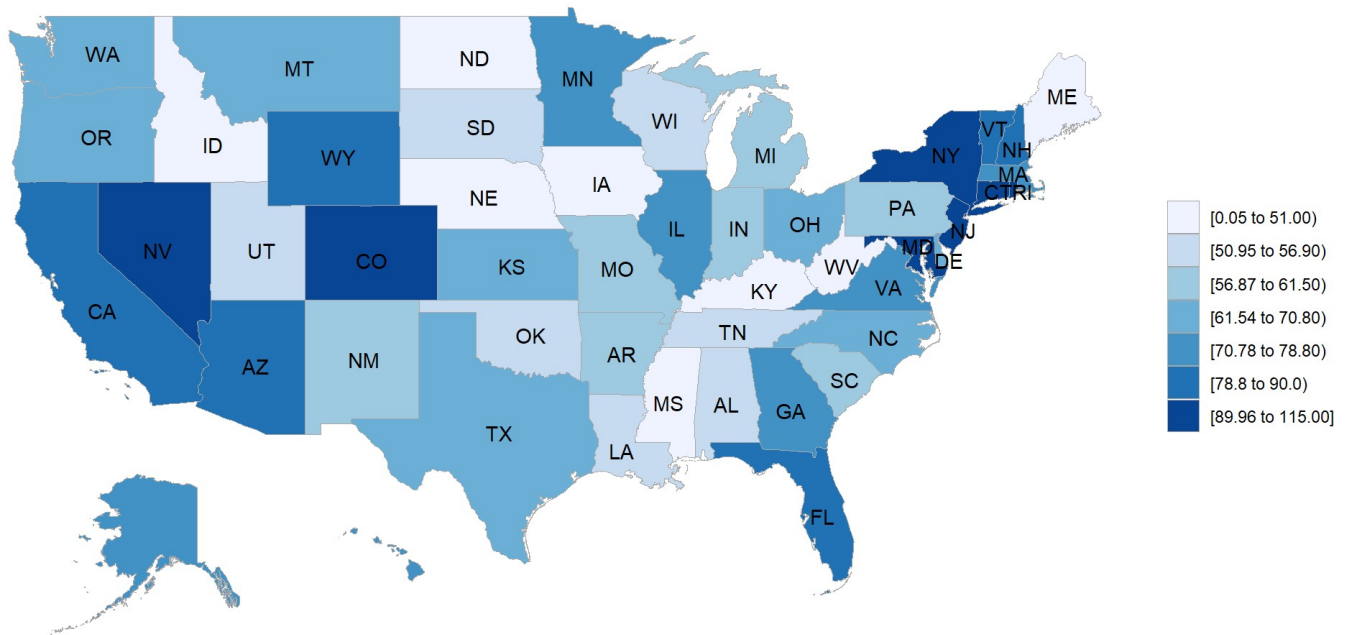
```
# loans per capita
```

```
lpc <- us_map[c("state", "loans_per_capita")]
```

```
colnames(lpc) <- c("region", "value")
```

```
state_choropleth(lpc, title = "Loans per 10,000 residents")
```

Loans per 10,000 residents



Connecticut residents took out, per capita, the second-highest number loans in the nation through LendingClub, a pioneer company in the “peer-to-peer” lending industry.

Between 2007 and 2015, Connecticut residents took out about 38 loans per 10,000 people.

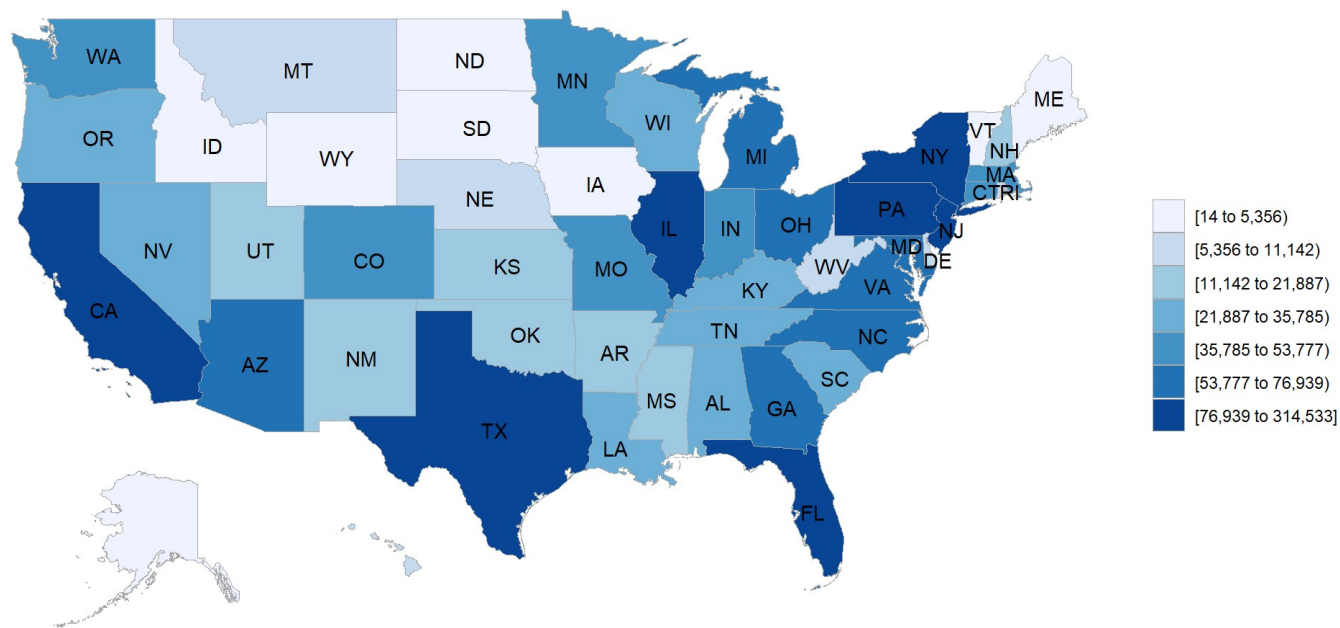
Connecticut’s rate of borrowing was a weak second to Nevada, where 44 loans per 10,000 people were issued.

```
l <- us_map[c("state", "loans")]
```

```
colnames(l) <- c("region", "value")
```

```
state_choropleth(l, title = "Total loans between 2007 and 2015")
```

Total loans between 2007 and 2015

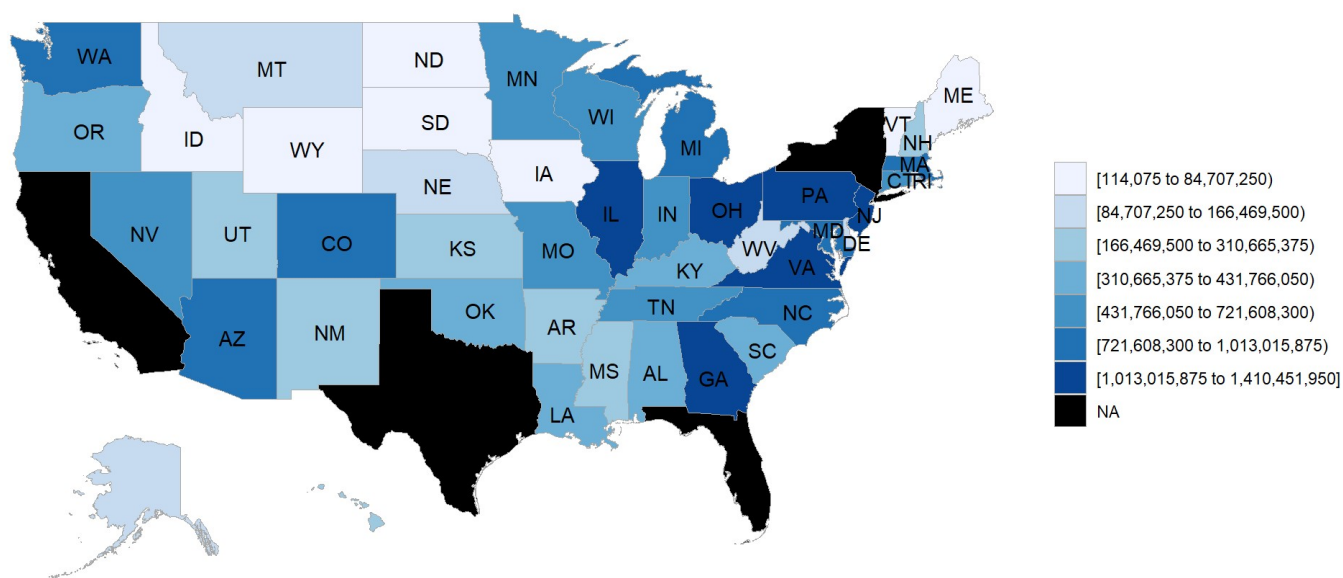


```
t <- us_map[c("state", "total")]
colnames(t) <- c("region", "value")

state_choropleth(t, title = "Total loan amounts between 2007 and 2015")
```

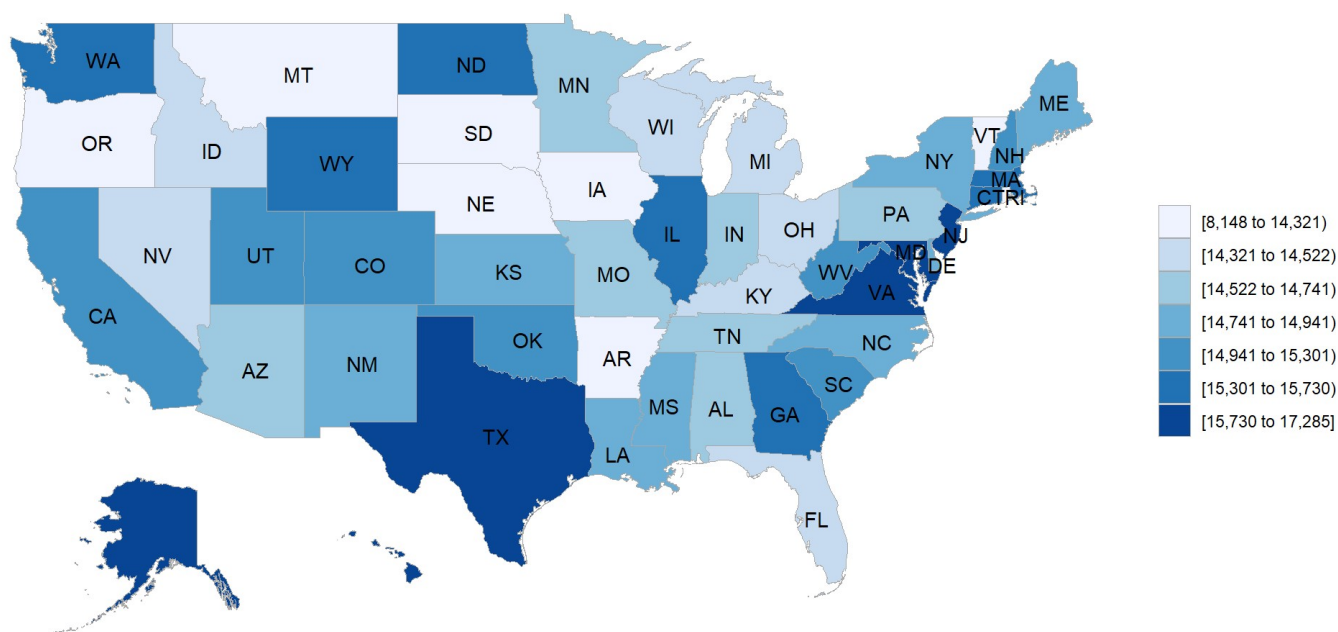
```
## Warning in self$bind(): The following regions were missing and are being
## set to NA: california, florida, new york, texas
```

Total loan amounts between 2007 and 2015



```
a <- us_map[c("state", "average")]  
colnames(a) <- c("region", "value")  
  
state_choropleth(a, title = "Average loan amounts between 2007 and 2015")
```

Average loan amounts between 2007 and 2015



Charting Connecticut data

```
ct_loans <- subset(loans, addr_state=="CT")

grade <- ct_loans %>%
  dplyr::group_by(grade) %>%
  dplyr::summarise(total=n()) %>%
  dplyr::mutate(percent=round(total/sum(total),2)) %>%
  dplyr::select(grade, percent) %>%
  dplyr::mutate(grade=factor(grade), grade=factor(grade, levels=rev(levels(grade))))

gg <- ggplot(grade, aes(y=grade, x=percent))
gg <- gg + geom_lollipop(point.colour="royalblue", point.size=3, horizontal=TRUE)
gg <- gg + scale_x_continuous(expand=c(0,0), labels=percent,
                             breaks=seq(0, 1, by=0.2), limits=c(0, .5))

#gg <- gg + coord_flip()
gg <- gg + labs(x=NULL, y=NULL,
               title="Loan grades in Connecticut",
               subtitle="Lending Club data between 2007 and 2015. Interest rates ran
ged from 5.32 percent to 28.49 percent,
               depending on the grade assigned to the loan from A to G.",
               caption="Source: Kaggle, LendingClub")
gg <- gg + theme_minimal(base_family="Arial Narrow")
gg <- gg + theme(panel.grid.major.y=element_blank())
gg <- gg + theme(panel.grid.minor=element_blank())
gg <- gg + theme(axis.line.y=element_line(color="#2b2b2b", size=0.15))
gg <- gg + theme(axis.text.y=element_text(margin=margin(r=-5, l=0)))
gg <- gg + theme(plot.margin=unit(rep(30, 4), "pt"))
gg <- gg + theme(plot.title=element_text(face="bold"))
gg <- gg + theme(plot.subtitle=element_text(margin=margin(b=10)))
gg <- gg + theme(plot.caption=element_text(size=8, margin=margin(t=10)))
gg
```

```
## Warning in grid.Call(C_stringMetric, as.graphicsAnnot(x$label)): font
## family not found in Windows font database

## Warning in grid.Call(C_stringMetric, as.graphicsAnnot(x$label)): font
## family not found in Windows font database

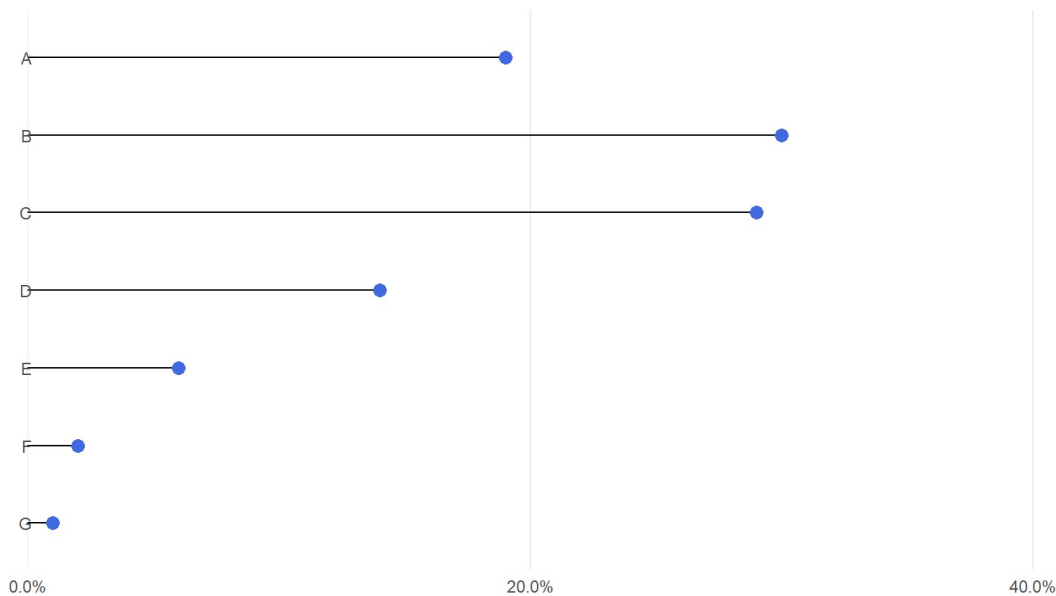
## Warning in grid.Call(C_stringMetric, as.graphicsAnnot(x$label)): font
## family not found in Windows font database

## Warning in grid.Call(C_stringMetric, as.graphicsAnnot(x$label)): font
## family not found in Windows font database
```

[illegible]

Loan grades in Connecticut

Lending Club data between 2007 and 2015. Interest rates ranged from 5.32 percent to 28.49 percent, depending on the grade assigned to the loan from A to G.



Source: Kaggle, LendingClub

About 29 percent of the loans issued via Lending Club were given a B grade. About 18 percent were ranked A.

```
purpose <- ct_loans %>%
  dplyr::group_by(purpose) %>%
  dplyr::summarise(total=n(), sum=sum(loan_amnt), average=mean(loan_amnt)) %>%
  dplyr::mutate(percent=round(total/sum(total)*100,2)) %>%
  dplyr::mutate(purpose=factor(purpose), purpose=factor(purpose, levels=rev(levels(purpose))))

gg <- ggplot(purpose, aes(y=purpose, x=total))
gg <- gg + geom_lollipop(point.colour="royalblue", point.size=3, horizontal=TRUE)
gg <- gg + scale_x_continuous(expand=c(0,0),
                             breaks=seq(0, 1, by=0.2), limits=c(0, 8000))

#gg <- gg + coord_flip()
gg <- gg + labs(x=NULL, y=NULL,
               title="Total loans issued in Connecticut",
               subtitle="Lending Club data between 2007 and 2015.",
               caption="Source: Kaggle, LendingClub")
gg <- gg + theme_minimal(base_family="Arial Narrow")
gg <- gg + theme(panel.grid.major.y=element_blank())
gg <- gg + theme(panel.grid.minor=element_blank())
gg <- gg + theme(axis.line.y=element_line(color="#2b2b2b", size=0.15))
gg <- gg + theme(axis.text.y=element_text(margin=margin(r=-5, l=0)))
gg <- gg + theme(plot.margin=unit(rep(30, 4), "pt"))
gg <- gg + theme(plot.title=element_text(face="bold"))
gg <- gg + theme(plot.subtitle=element_text(margin=margin(b=10)))
gg <- gg + theme(plot.caption=element_text(size=8, margin=margin(t=10)))
gg
```

```
## Warning: Removed 2 rows containing missing values (geom_lollipop).
```

```
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## font family not found in Windows font database
```

```
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## font family not found in Windows font database
```

```
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## font family not found in Windows font database
```

```
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## font family not found in Windows font database
```

```
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## font family not found in Windows font database
```

```
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## font family not found in Windows font database
```

```
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## font family not found in Windows font database
```

```
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## font family not found in Windows font database
```

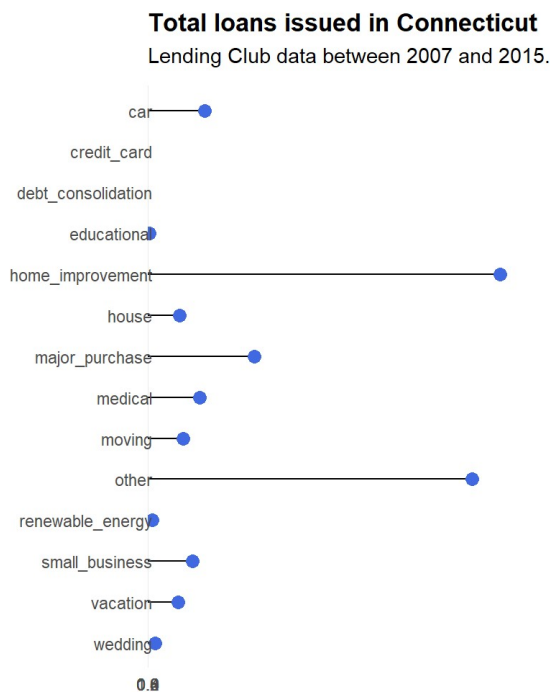
```
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## font family not found in Windows font database
```

```
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## font family not found in Windows font database
```

```
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## font family not found in Windows font database
```

```
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## font family not found in Windows font database
```

```
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## font family not found in Windows font database
```

Source: Kaggle, LendingClub

More than 10,000 of the loans in Connecticut were for the purposes of debt consolidation or paying off credit card debt.

In contrast, during the same period, only 70 loans were for vacations and 50 to help fund weddings.

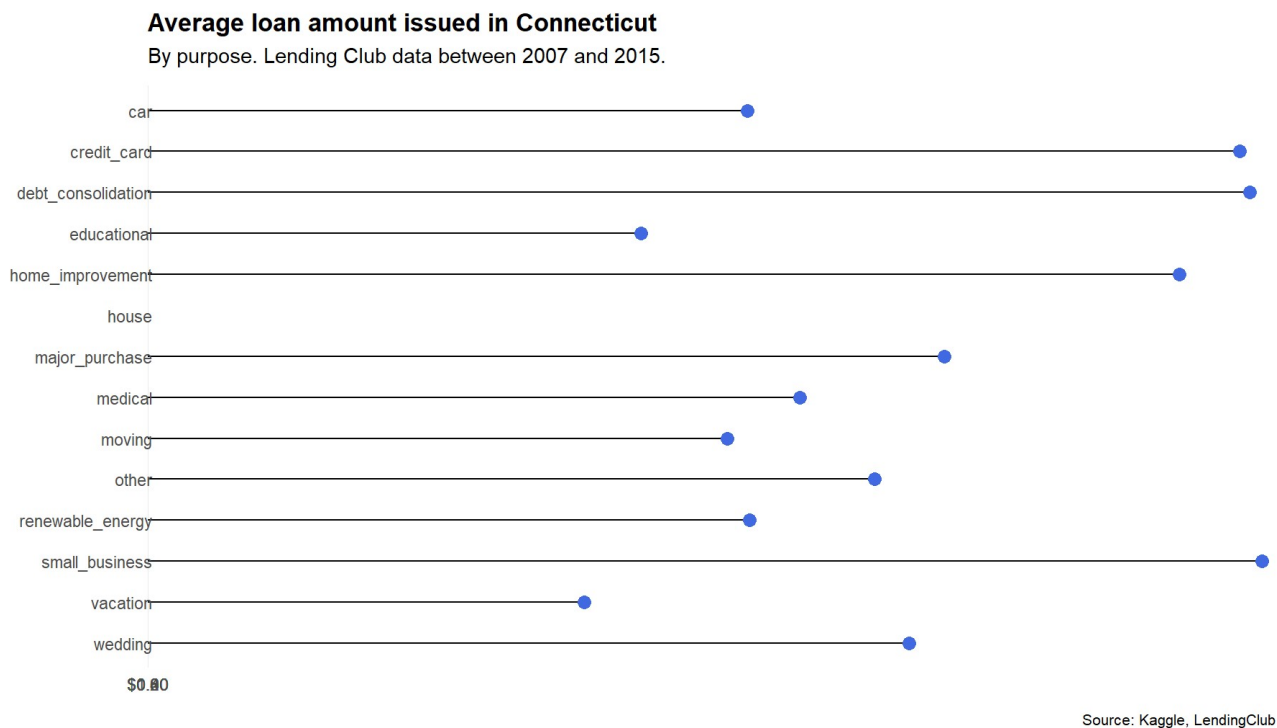
```
## Avg
gg <- ggplot(purpose, aes(y=purpose, x=average))
gg <- gg + geom_lollipop(point.colour="royalblue", point.size=3, horizontal=TRUE)
gg <- gg + scale_x_continuous(expand=c(0,0), labels=dollar,
                             breaks=seq(0, 1, by=0.2), limits=c(0, 16500))

#gg <- gg + coord_flip()
gg <- gg + labs(x=NULL, y=NULL,
               title="Average loan amount issued in Connecticut",
               subtitle="By purpose. Lending Club data between 2007 and 2015.",
               caption="Source: Kaggle, LendingClub")

gg <- gg + theme_minimal(base_family="Arial Narrow")
gg <- gg + theme(panel.grid.major.y=element_blank())
gg <- gg + theme(panel.grid.minor=element_blank())
gg <- gg + theme(axis.line.y=element_line(color="#2b2b2b", size=0.15))
gg <- gg + theme(axis.text.y=element_text(margin=margin(r=-5, l=0)))
gg <- gg + theme(plot.margin=unit(rep(30, 4), "pt"))
gg <- gg + theme(plot.title=element_text(face="bold"))
gg <- gg + theme(plot.subtitle=element_text(margin=margin(b=10)))
gg <- gg + theme(plot.caption=element_text(size=8, margin=margin(t=10)))
gg
```

```
## Warning: Removed 1 rows containing missing values (geom_lollipop).
```

[illegible]



Seven loans were requested to help pay for school tuition, while nearly 900 loans were filed for home improvements such as kitchen upgrades or roof repairs. More than \$11 million in loans were issued between 2007 and 2015 to assist with home improvements—the most amount after credit card and debt consolidation.

More than 70 residents of Connecticut were issued loans to fund their vacation trips to Europe or Florida. Nearly 50 were issued to pay for weddings—the borrower's own or a daughter's—according to the description fields. The average loan was for \$11,000.

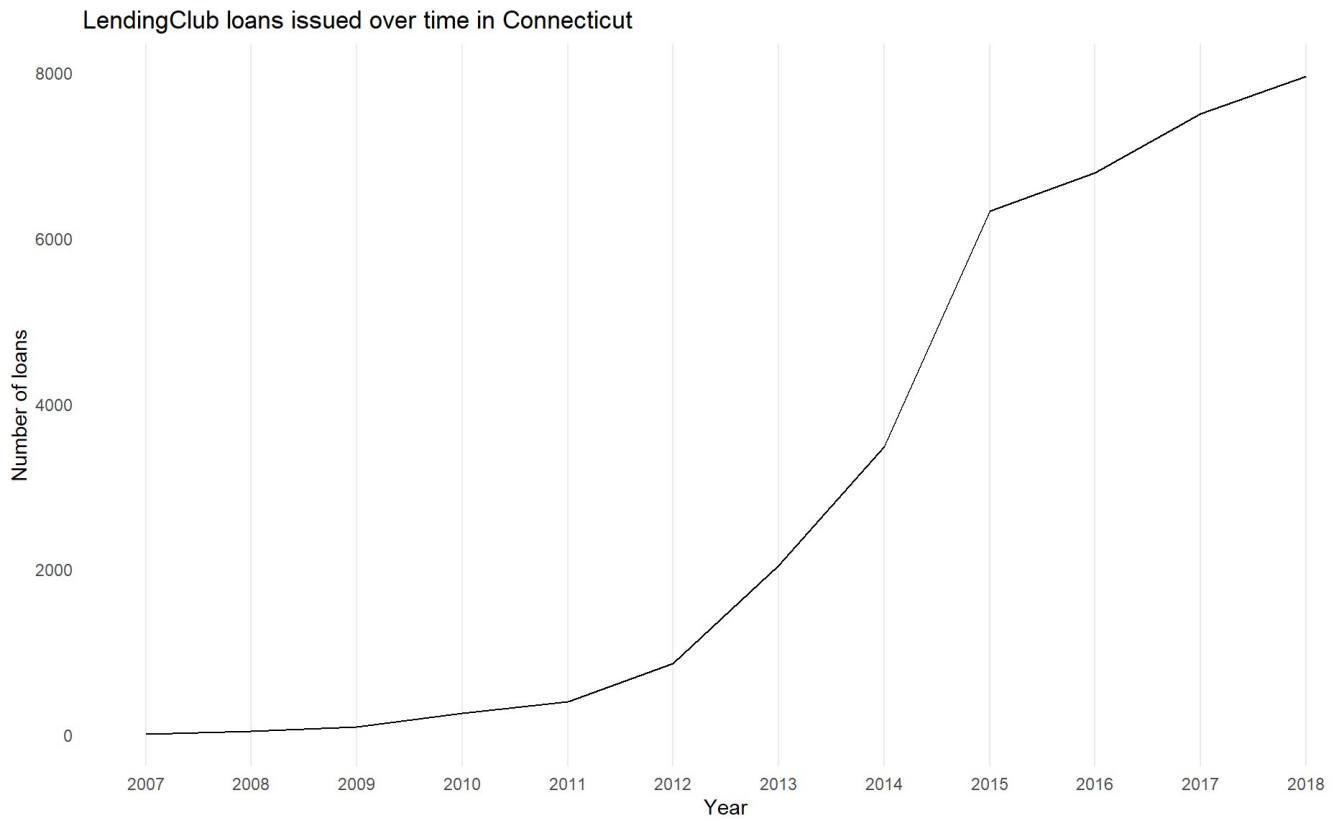
Annual growth

```
ct_loans$year <- gsub(".*-", "", ct_loans$issue_d)

years <- ct_loans %>%
  dplyr::group_by(year) %>%
  dplyr::summarise(loans=n())

ggplot(years, aes(x=year, y=loans, group=1)) + geom_line() +
  labs(x = "Year", y = "Number of loans",
       title = "LendingClub loans issued over time in Connecticut") +
  theme_minimal(base_family="Arial Narrow") +
  theme(panel.grid.major.y=element_blank()) +
  theme(panel.grid.minor=element_blank())
```

```
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## font family not found in Windows font database
```



The number of people in Connecticut using LendingClub loans has been increasing since 2007— with 13,500 loans issued as of 2015. The highest point was last year with more than 6,000 loans issued.