# Interactive Maps for Dummies:

## Lecture 1: ggplot2 and draw maps

SHI Shuang

The Univeristy of Hong Kong

Visiting Research Fellow at Brown University

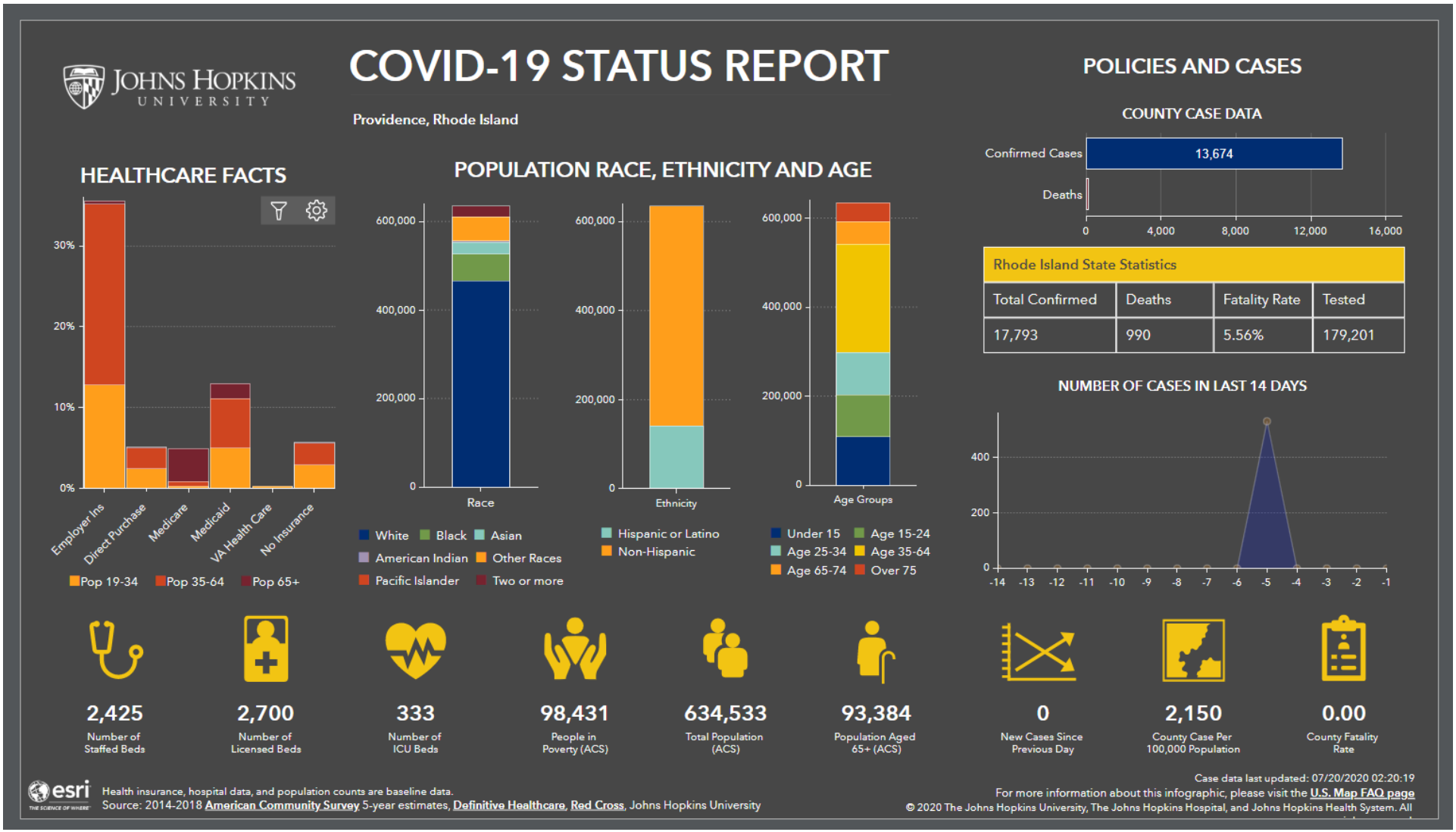
Jul. 21, 2020

Prepared for HKBU Zoom Lectures

# Today

- R basics
- ggplot2
- Introduction to the spatial world

# Plot A Simple Bar Graph

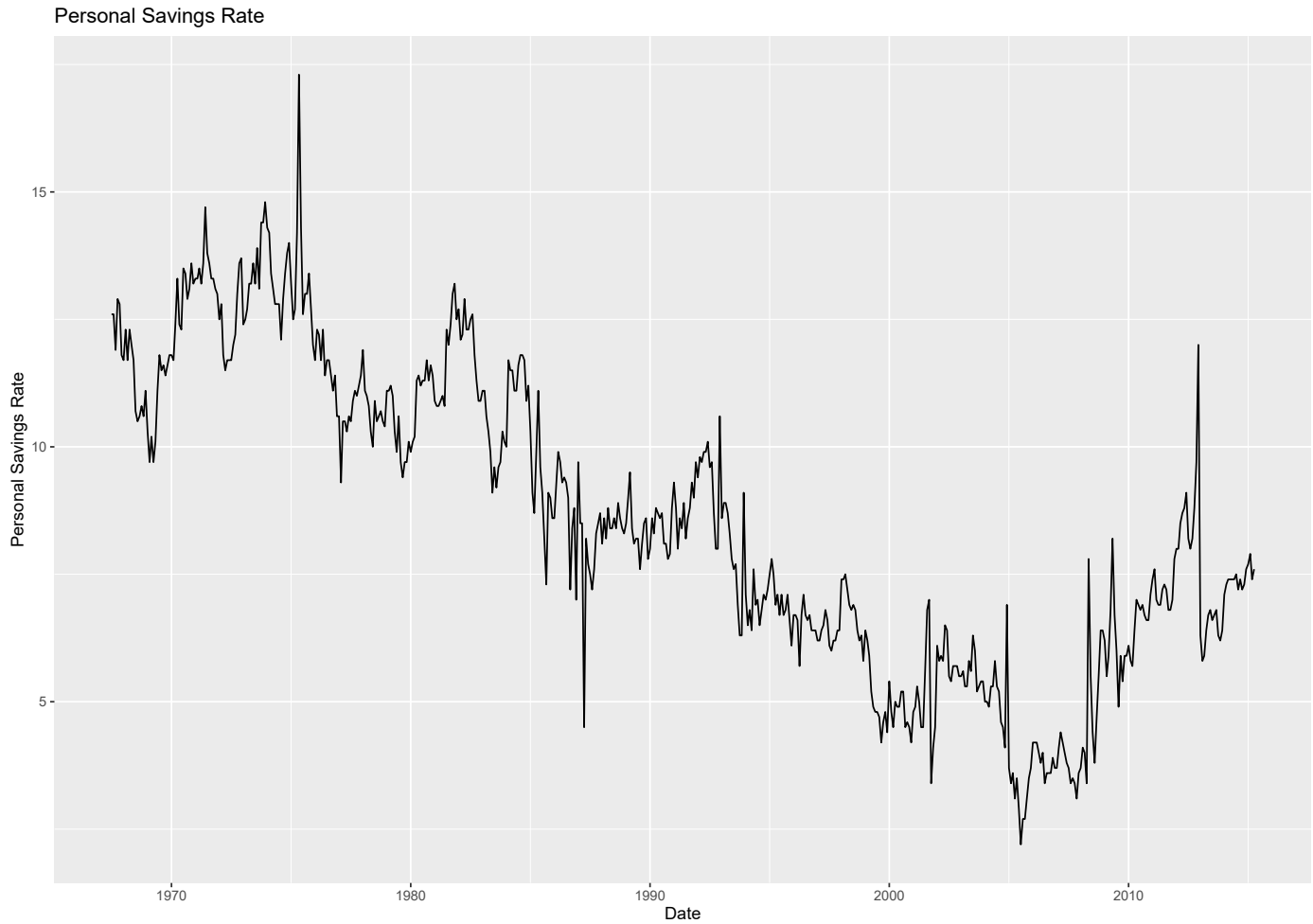


# Plot A Simple Line Graph

- The most common time-dependent graph is the time series line graph
  - We use the Economics time series that come with the *ggplot2*, which contains US monthly economic data from Jan. 1967 to Jan. 2015
  - We first try to plot the personal saving rate data （*psavert*）

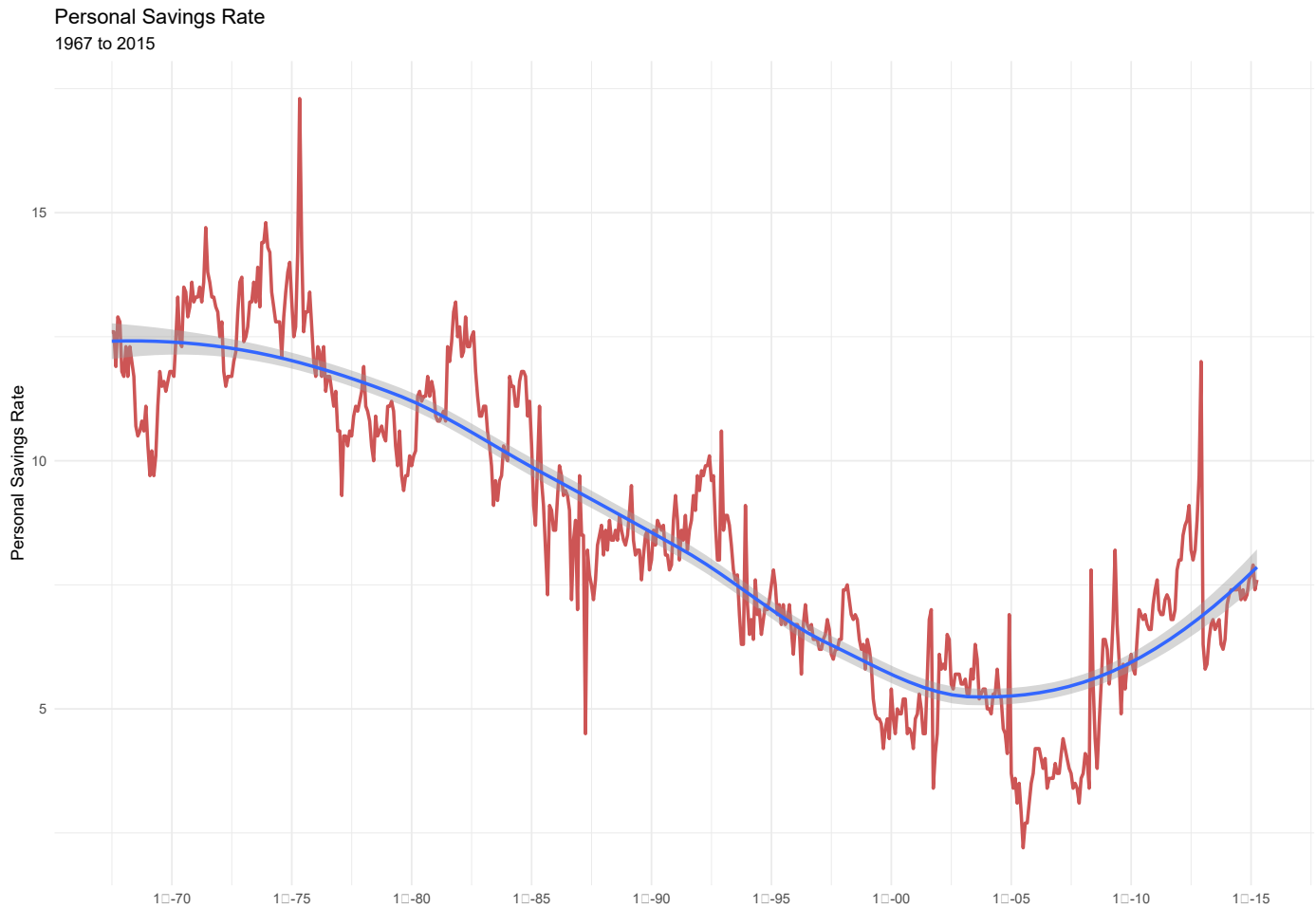
```
 2  library(ggplot2)
 3  # Plot time series data
 4  ## using the economics data comes with ggplot2 library
 5  economics <- economics
 6
 7  ## Simple line plot
 8  ggplot(economics, aes(x = date, y = psavert)) +
 9    geom_line() +
10    labs(title = "Personal Savings Rate",
11         x = "Date",
12         y = "Personal Savings Rate")
```

# Plot A Simple Line Graph



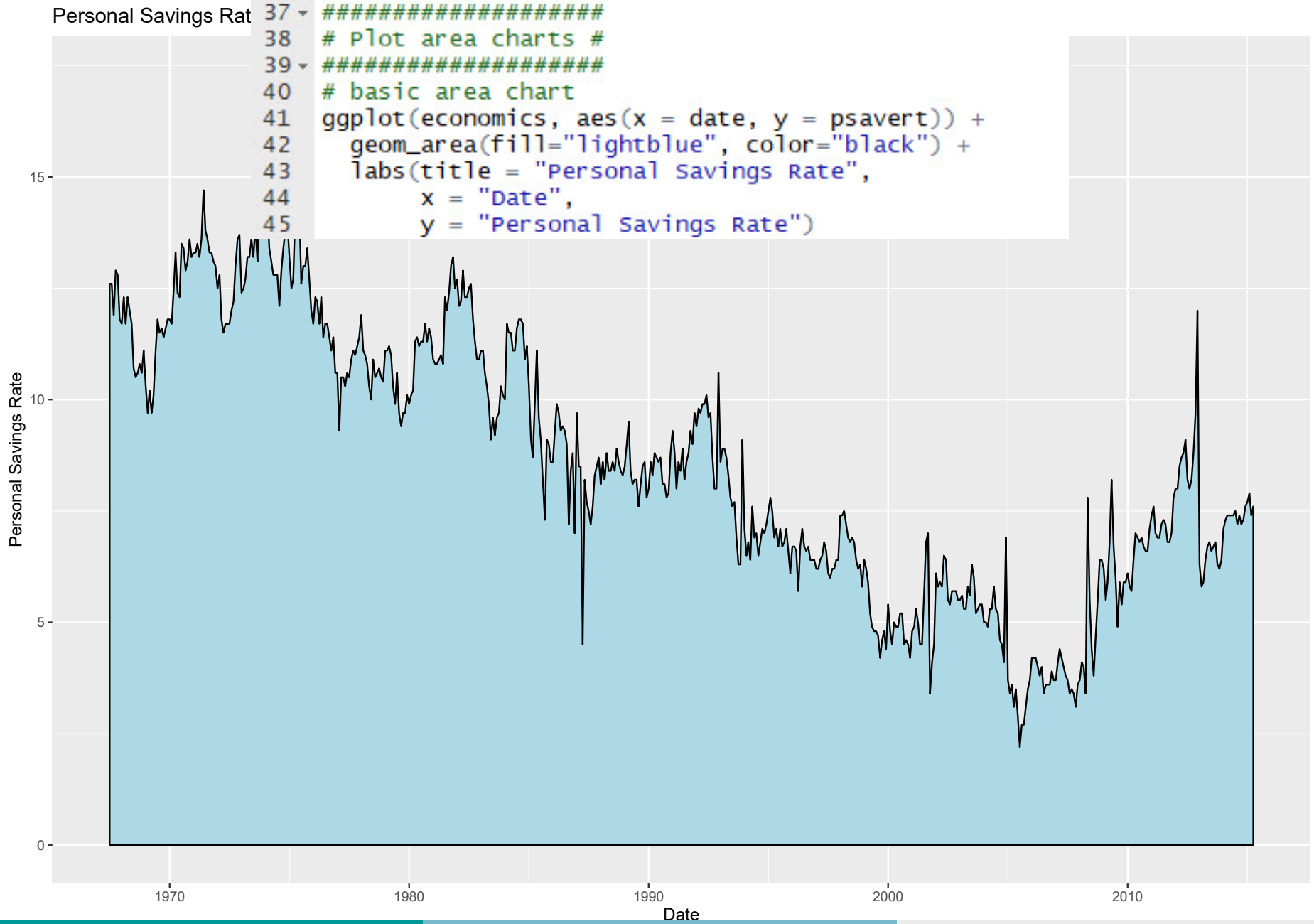
# Make it prettier!

- Make it prettier by adjusting the scale (*scale_x_date()*) and theme (*theme_minimal()*)
- Also add the smooth line using *geom_smooth* to show the trend



# Can Use an Area Chart instead

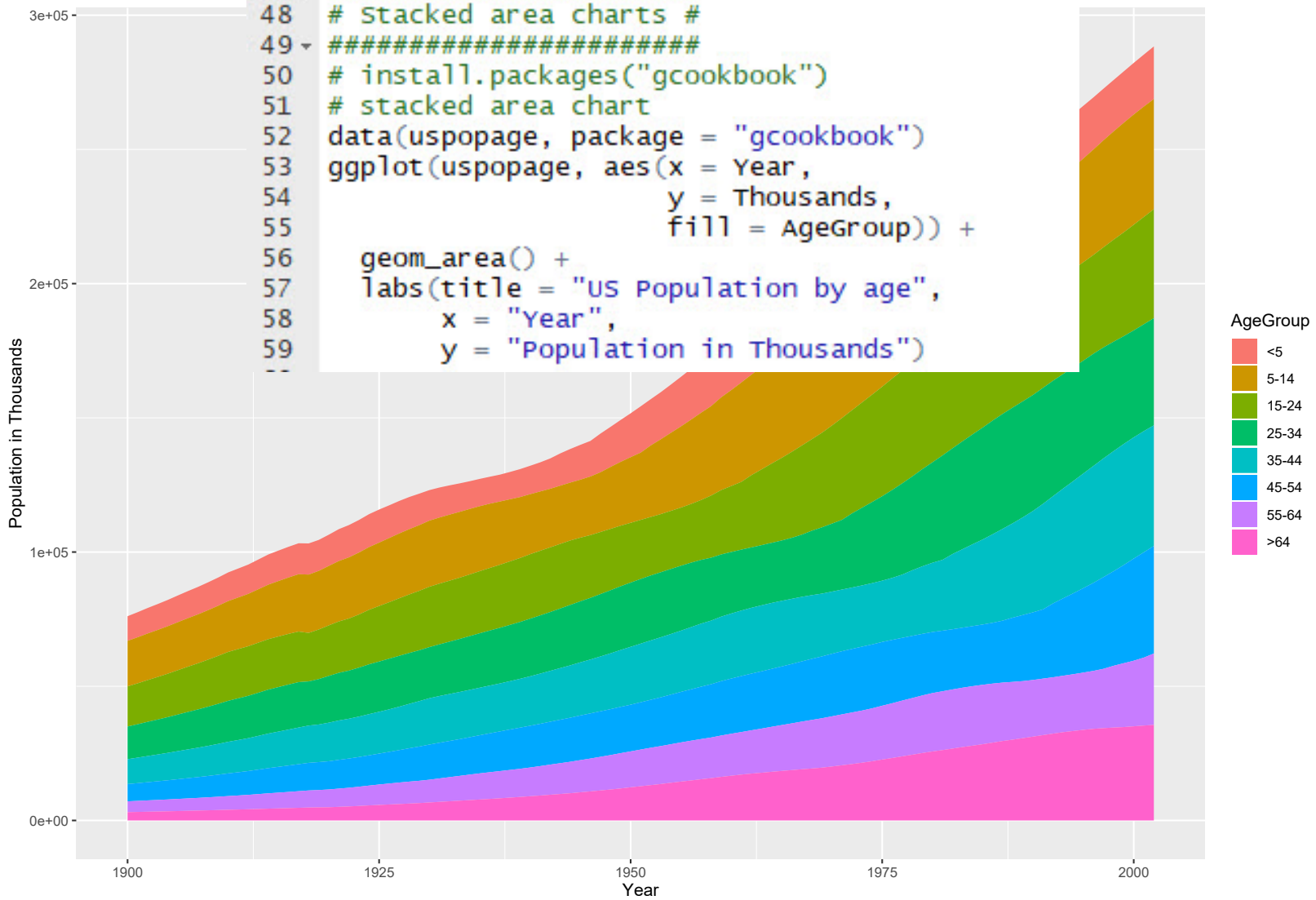
```
37 ####################
38 # Plot area charts #
39 ####################
40 # basic area chart
41 ggplot(economics, aes(x = date, y = psavert)) +
42   geom_area(fill="lightblue", color="black") +
43   labs(title = "Personal Savings Rate",
44        x = "Date",
45        y = "Personal Savings Rate")
```



# A Stacked Area Chart

- Use a stacked area chart to show differences between groups over time

```
47 ########################
48 # Stacked area charts #
49 ########################
50 # install.packages("gcookbook")
51 # stacked area chart
52 data(uspopage, package = "gcookbook")
53 ggplot(uspopage, aes(x = Year,
54                      y = Thousands,
55                      fill = AgeGroup)) +
56   geom_area() +
57   labs(title = "US Population by age",
58        x = "Year",
59        y = "Population in Thousands")
```



# Plot Multivariate Time Series: Prepare Data

- *quantmod* (Quantitative Financial Modelling & Trading Framework for R) is a powerful library to retrive and analyze financial data (http://www.quantmod.com/)
  - The *getSymbols()* function can be used to retrive financial data from Internet
  - Supported online financial database including yahoo, google, FRED, Oanda etc.

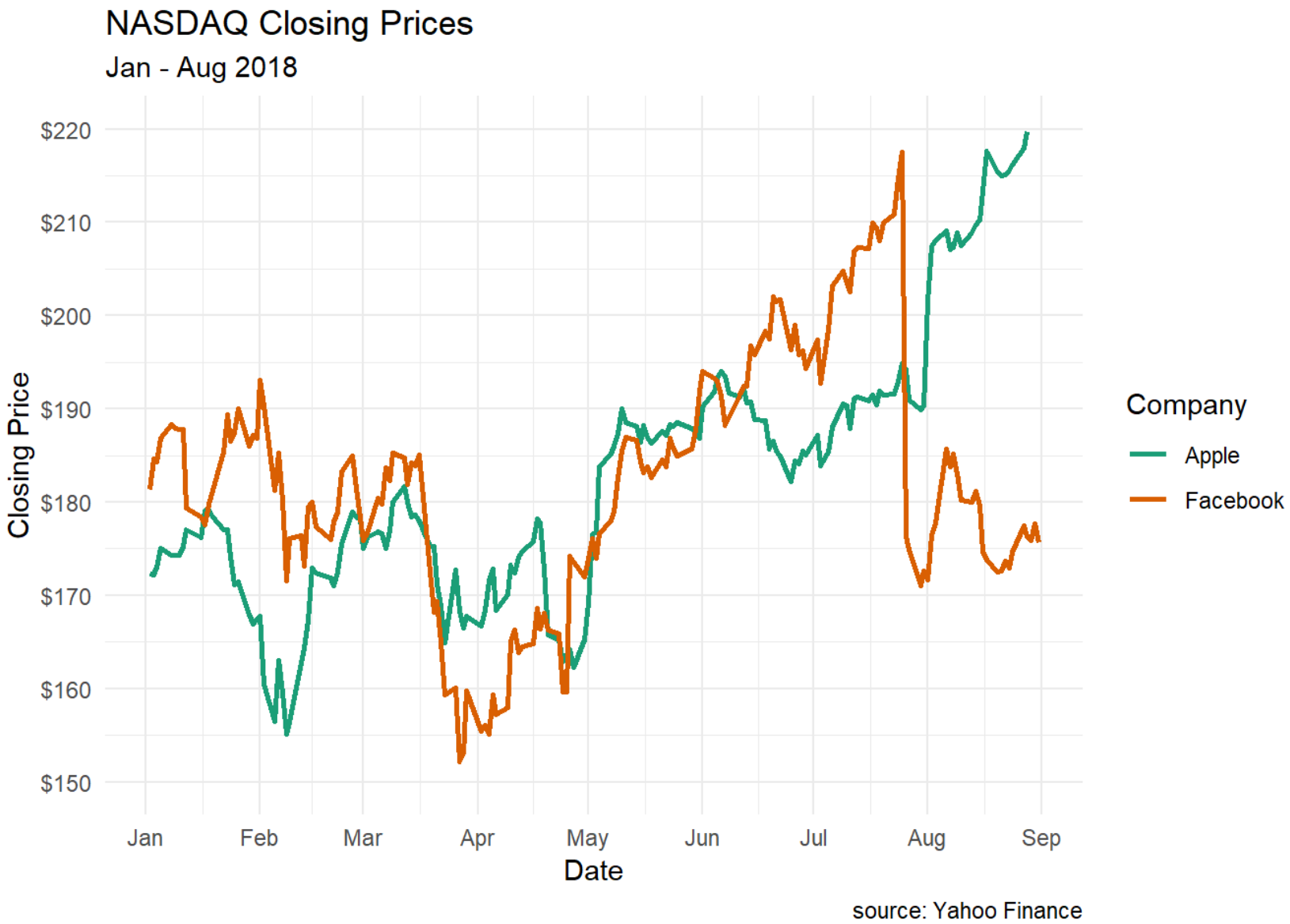
```
31  library(quantmod)
32  library(dplyr)
33
34  ## get apple (AAPL) closing prices
35  apple <- getSymbols("AAPL",
36                      return.class = "data.frame",
37                      from="2018-01-01")
38
39  apple <- AAPL %>%
40    mutate(Date = as.Date(row.names(.))) %>%
41    select(Date, AAPL.Close) %>%
42    rename(Close = AAPL.Close) %>%
43    mutate(Company = "Apple")
```

# Plot Multivariate Time Series

- Compare the stock prices between Apple and Facebook
  - Need to append the data using the *rbind* function, and use color to facet the data

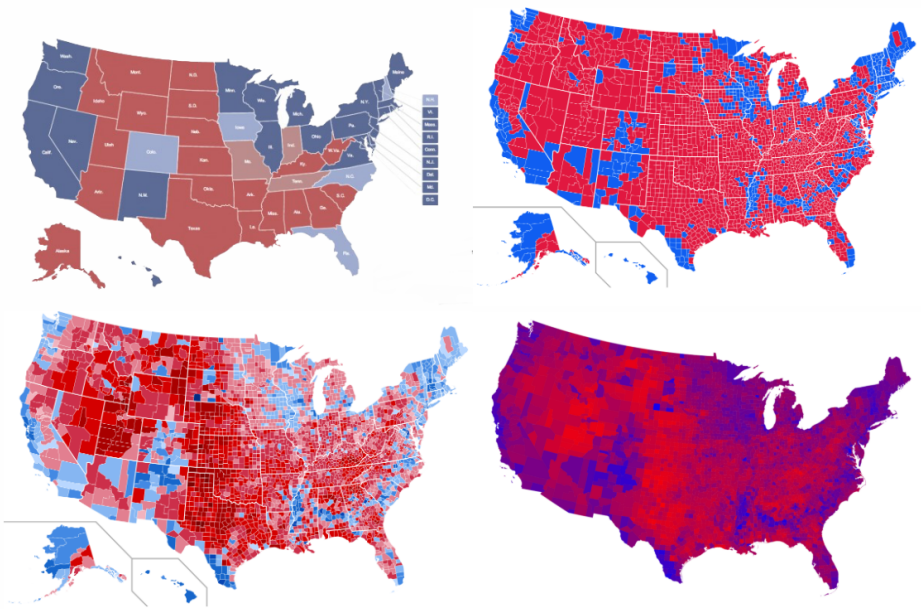
```
56  ## Append data for both companies
57  mseries <- rbind(apple, facebook)
58
59  ## plot data
60  ggplot(mseries,
61         aes(x=Date, y= Close, color=Company)) +
62    geom_line(size=1) +
63    scale_x_date(date_breaks = '1 month',
64                 labels = scales::date_format("%b")) +
65    scale_y_continuous(limits = c(150, 220),
66                       breaks = seq(150, 220, 10),
67                       labels = scales::dollar) +
68    labs(title = "NASDAQ Closing Prices",
69         subtitle = "Jan - Aug 2018",
70         caption = "source: Yahoo Finance",
71         y = "Closing Price") +
72    theme_minimal() +
73    scale_color_brewer(palette = "Dark2")
```

# Plot Multivariate Time Series (Cont'd)



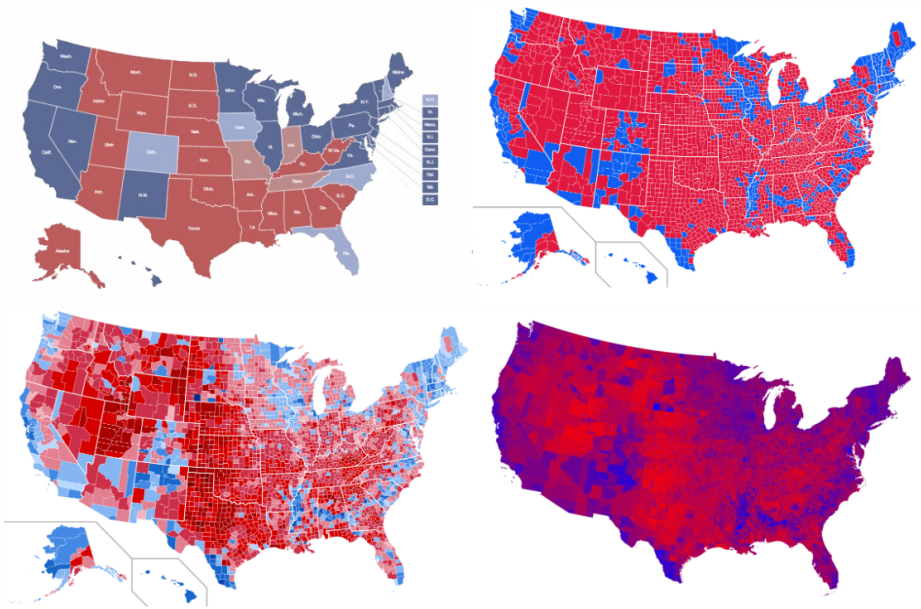
# Draw Map

- Choropleth maps show geographical regions colored, shaded, or graded according to some variable
- *R* is not as powerful as ArcGIS with GIS data, but *ggplot2* can draw map
- Before draw a map, decide three things: 1) geographic unit, 2) classification, 3) color or shape or dot density



# Draw Map

- Choropleth maps show geographical regions colored, shaded, or graded according to some variable
- *R* is not as powerful as ArcGIS with GIS data, but *ggplot2* can draw map
- Before draw a map, decide three things: 1) geographic unit, 2) classification, 3) color or shape or dot density



# Boundary map

- First step: drawing boundaries
- Either by using the maps package loaded with boundary data or bringing in external data in the form of shapefiles
- Several libraries could draw map, we will use **ggmap**, and **rgdal**
  - **ggmap** visualize spatial data and models on top of static maps from various online sources (e.g Google Maps and Stamen Maps)

# Shapefiles

- Shapefile is a file format specifically for geographic data
- Encodes points, lines, and polygons in geographic space
- File extension is .shp
- Other files accompany the .shp file with extensions .dbf and .prj
  - .dbf contains attribute format of shapefile
  - .prg contains information about the projection of the coordinates
- You could load in shapefiles to R using **rgdal**
- e.g. Downloading hk district map from `http://opendata.esrichina.hk/datasets/eea8ff2f12b145f7b33c4eef4f045513_0/data`

# Map Projections

- Mapping compromise accuracy, readability, and aesthetics
- To represent a three-dimensional thing (the earth) with two dimensions on the screen or paper
- Projections is the math beyond placing three-dimensional space on a two-dimensional surface
- Each projection distorts size of area, distance in different way
- Many of the projections have parameters for latitude and longitude

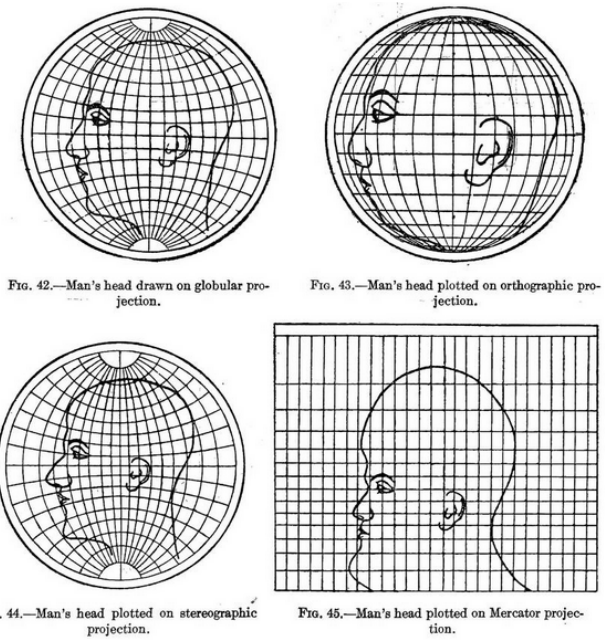


FIG. 42.—Man's head drawn on globular projection.

FIG. 43.—Man's head plotted on orthographic projection.

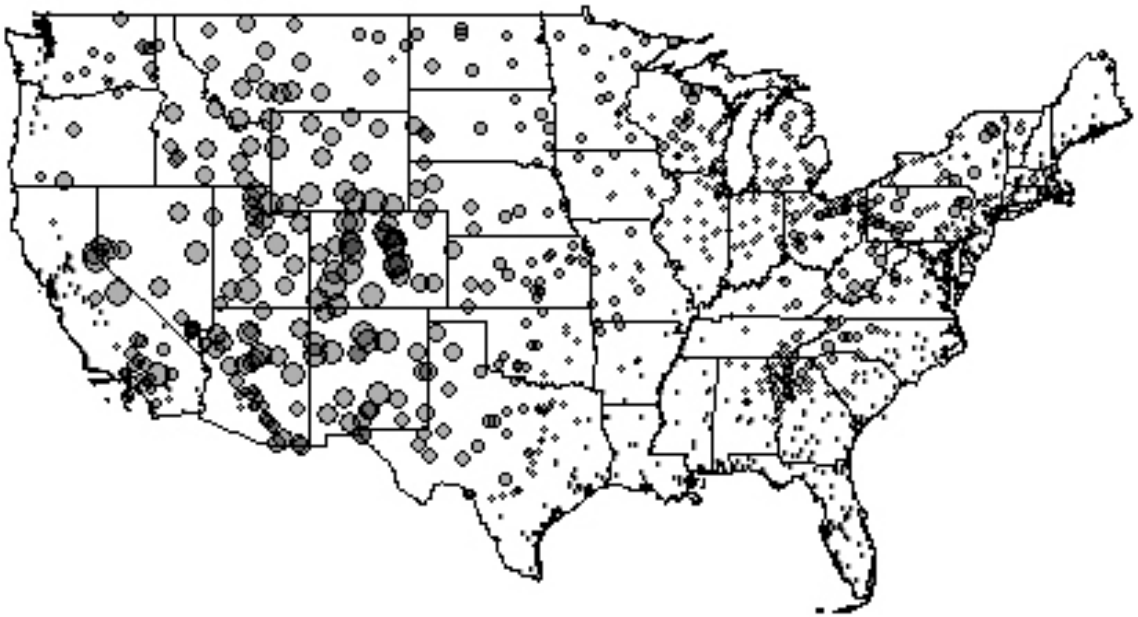
FIG. 44.—Man's head plotted on stereographic projection.

FIG. 45.—Man's head plotted on Mercator projection.

# Draw Locations Using an Imported Shapefile

# Draw Locations Using an Imported Shapefile

- What if we do not project?



# Your Turn

- Download the Starbucks data from `https://community.periscopedata.com/t/80fyna/starbucks-locations`
- And Hong Kong 18 district shapefile from `http://opendata.esrichina.hk/datasets/eea8ff2f12b145f7b33c4eef4f045513_0/data`
- Select the Starbucks for Hong Kong based on **Country==”CN” & Country.Subdivision==”91”**
- Draw the locations for all the Starbucks in Hong Kong