

# Statistics with Recitation: TA Session

Danny Po-Hsien Kang (康柏賢)

September 9, 2025

# TA Session's Website

- Now you can find everything on [TA's website](#).
  - Slides, In-Class Practices, Answers...



# Today's agenda

- 1 Read Data
  - `getwd()`, `setwd()`
  - `read.csv()`
- 2 Draw Graphs
  - `geom_point()`
  - `geom_histogram()`
  - `geom_boxplot()`
  - `geom_dotplot()`

## Before today's lecture

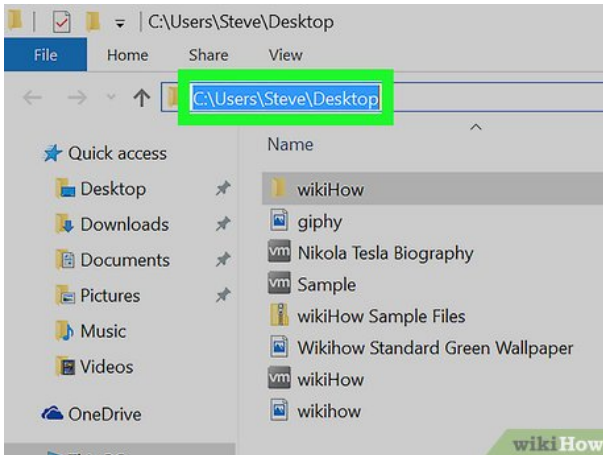
- Please make sure you have downloaded the following dataset from the link on NTU COOL<sup>1</sup>.
  - county.csv
  - loan50.csv
- We'll learn how to analyze datasets with many different datasets, so please download them through the same link each time.
- Besides, please create a folder/directory on your computer to put all the datasets used in TA sessions.

---

<sup>1</sup>You can also download the same dataset from [OpenIntro Website](#), which provides the complete information about the dataset.

# Path of a file

- A **path** for a file is the place where you save it.



# Get/Set Current Directory

- `getwd()`: Get current working directory

```
getwd() #Example Output: "C:/your/working/directory"
```

- `setwd()`: Change your working directory<sup>2</sup>

```
setwd("C:\\your\\working\\directory") #Error!  
setwd("C:\\\\your\\\\working\\\\directory") #Works on Windows!  
setwd("C:/your/working/directory") #Works everywhere!
```

---

<sup>2</sup>You can use `dir.exists()` to check if the directory exists.

## Supplement: Escape Character \

- If you directly copy the path in Windows (C:/your/working/directory) and put it in `setwd()`, you'll get an error. But why?
  - Because “\” is an escape character in R!
- An **escape character** is used to represent special characters within a string.
  - `\n`: newline character
  - `\t`: horizontal tab character
- This is why the error message shows: `'\U' used without ...!`
- If you want to type `\` in R, use `\\`
  - Similar for `"`, you need to use `\"`

# Read CSV File

- **Syntax:**

```
df <- read.csv("file.csv") #Work if file in working directory  
df <- read.csv("C:/your/working/directory/file.csv")
```

- **Example:**

```
county <- read.csv("county.csv")  
loan <- read.csv("loan50.csv")
```



## Supplement: Using ~ and .. as directory

- In R, there is a personal base folder called **home directory**
  - You can check your home directory by `path.expand("~/")`.
  - Won't change by applying `setwd()`
  - Can be changed by using `Sys.setenv` or options in the Global Options.
- ~ is a shorthand path notation, representing the home directory.
  - You may write: `data <- read.csv("~/Documents/mydata.csv")`
- .. refers to “one level up” from the current working directory.
  - Can be chained: `../..` means “go up two folders”.
  - `read.csv("../somefile.csv")` reads the csv from the parent folder of the current working directory.

# What's Recorded in Dataset?

- `county.csv`
  - **county**: County name in United States
  - **homeownership**: Home ownership rate, 2006-2010
  - **multi\_unit**: Percent of housing units in multi-unit structures, 2006-2010
- `loan50.csv`
  - **annual\_income**: Annual income of the applicant
  - **interest\_rate**: Interest rate of the loan the applicant received
  - **grade**: Grade associated with the loan.

# Scatter Plot: `geom_point()`

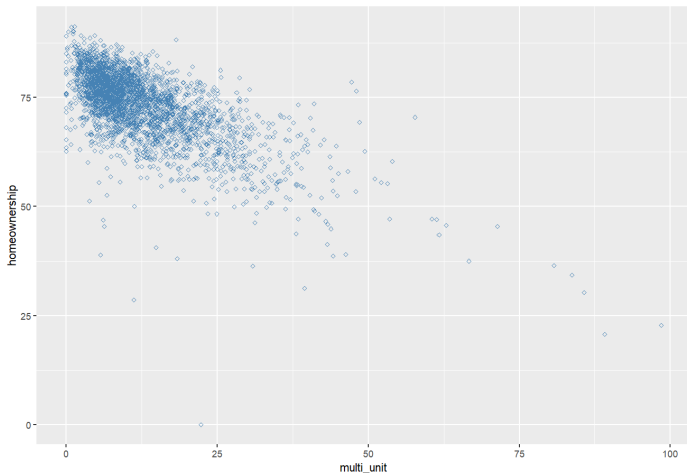


Figure: Scatter plot of multiple unit ratio and home ownerships.

# Scatter Plot: `geom_point()`

- **Syntax:**

```
ggplot(data, aes(x = ..., y = ...)) +  
  geom_point(color = ...,  
             size = ...,  
             shape = ...  
            )
```

- **Example:**<sup>34</sup>

```
ggplot(county, aes(x = multi_unit, y = homeownership)) +  
  geom_point(color = "steelblue",  
            size = 1,  
            shape = 5  
            )
```

---

<sup>3</sup>R color [cheatsheet](#).

<sup>4</sup>ggplot2 [Quick Reference](#): shape

# Histogram: `geom_histogram()`

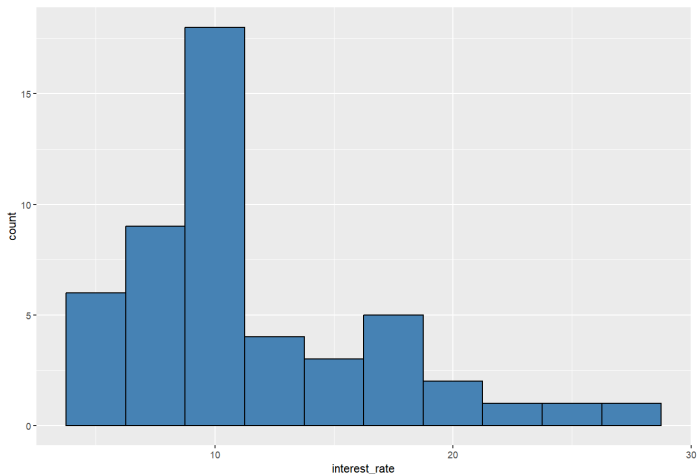


Figure: Histogram of interest rates.

# Histogram: geom\_histogram()

- **Syntax:**

```
ggplot(data, aes(x = ...)) +  
  geom_histogram(binwidth = ...,  
                 fill = ...,  
                 color = ...  
                )
```

- **Example:**

```
ggplot(loan, aes(x = interest_rate)) +  
  geom_histogram(binwidth = 2.5,  
                 fill = "steelblue",  
                 color = "black"  
                )
```

## Boxplot: `geom_boxplot()`

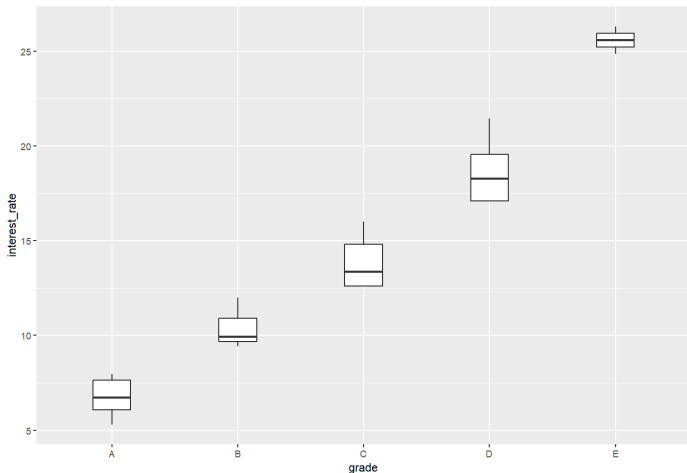


Figure: Box plot of interest rates by verified income.

# Boxplot: geom\_boxplot()

- **Syntax:**

```
ggplot(data, aes(x = ..., y = ...)) +  
  geom_boxplot(width = ...) #optional
```

- **Example:**

```
ggplot(loan, aes(x = grade, y = interest_rate)) +  
  geom_boxplot(width = 0.3)
```



## Dotplot: geom\_dotplot()

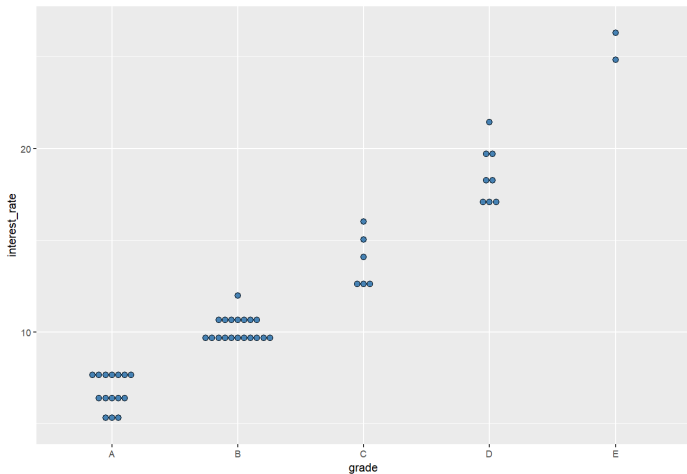


Figure: Dot plot of interest rates by verified income.

# Dotplot: geom\_dotplot()

- **Syntax:**

```
ggplot(data, aes(x = ..., y = ...)) +  
  geom_dotplot(binaxis = ..., # "x" (default) or "y"  
               stackdir = ..., # "up", "down", "center", ...  
               dotsize = ...,  
               fill = ...  
            )
```

- **Example:**

```
ggplot(loan, aes(x = "grade", y = interest_rate)) +  
  geom_dotplot(binaxis = 'y',  
               stackdir = 'center',  
               dotsize = 0.5,  
               fill = "lightblue"  
            )
```