# Statistics with Recitation: TA Session

Danny Po-Hsien Kang (康柏賢)

October 7, 2025

# Today's agenda

1. Paste strings together: `paste()`
   - `paste()`
   - `paste0()`

2. Date Type: Date
   - `as.Date()`

3. Draw Graphs
   - `geom_line()`

4. About the R Quiz
   - Reminders
   - Example Questions

## Today's Dataset

- Please download the following dataset:
    - ames.csv

  from the TA Session's Website or the OpenIntro website.
- The dataset includes the housing prices in Ames, Iowa, from 2006 to 2010.
- Please import the following datasets into RStudio.

```
housing <- read.csv("data/ames.csv")
```

# Paste strings together: paste()

- **Syntax:**

```
paste(vec_1, vec_2, sep = ..., collapse = ..., )
```

- **Example:**

```
paste("test", 1:5, sep = "_") # default = " "
[1] "test_1" "test_2" "test_3" "test_4" "test_5"

paste(c("A","B","C"), collapse = ", ")
[1] "A, B, C"

paste("test", 1:5, sep = "^", collapse = "+")
[1] "test^1+test^2+test^3+test^4+test^5"
```

# Paste strings together: `paste()`

- `paste()` works pairwise across inputs; shorter vectors recycle:

```
paste(c("A", "B", "C", "D"), 1:2, sep = "_")
[1] "A_1" "B_2" "C_1" "D_2"
```

- You can combine more than two vectors or strings.

```
paste("ID", 1:3, "Section", c("A", "B"), sep = "_")
[1] "ID_1_Section_A" "ID_2_Section_B" "ID_3_Section_A"
```

# Paste strings together: paste0()

- paste0() is a shorthand for paste(..., sep = "").
    - So no need to assign sep = ...!
- **Syntax:**

```
paste0(vec_1, vec_2, collapse = ..., )
```

- **Example:**

```
paste0("test", 1:5)
[1] "test1" "test2" "test3" "test4" "test5"

paste0(c("A","B","C"), collapse = ", ")
[1] "A, B, C"

paste0("test", 1:5, collapse = "+")
[1] "test1+test2+test3+test4+test5"
```

# Turn strings into date type: `as.Date()`

- Sometimes we hope date data can keep the "numeric" properties instead of being a character.
  - `as.Date()` will turn a specific type of string into `Date` data.
  - It might show `unknown` when you view in the dataframe though . . . .
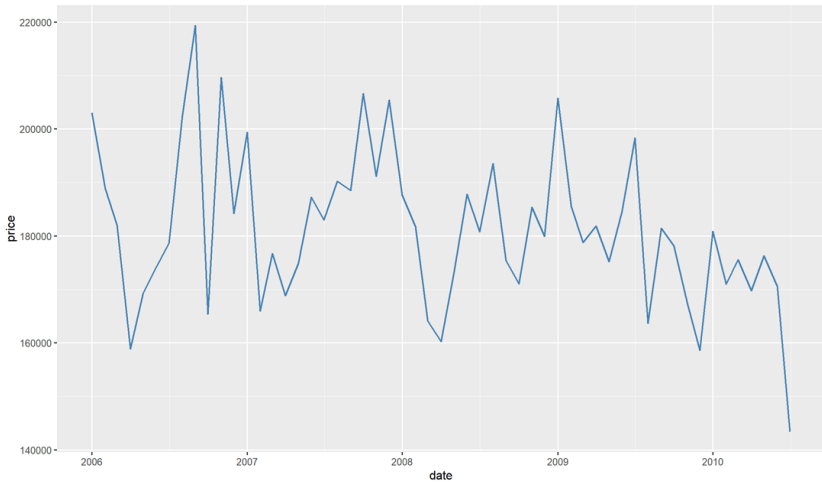
- **Syntax:**

```
as.Date(string)
as.Date(string, format = "%m-%d-%Y") #default %Y-%m-%d
```

- **Example:**

```
as.Date("2025-01-02")
[1] "2025-01-02"

as.Date("01-02-2025", format = "%m-%d-%Y")
[1] "2025-01-02"
```

# Draw Line Graph: geom_line()

# Draw Line Graph: geom_line()

- **Syntax:**

```
ggplot(data, aes(x = ..., y = ...)) +
  geom_line(linewidth = ..., color = ...)
```

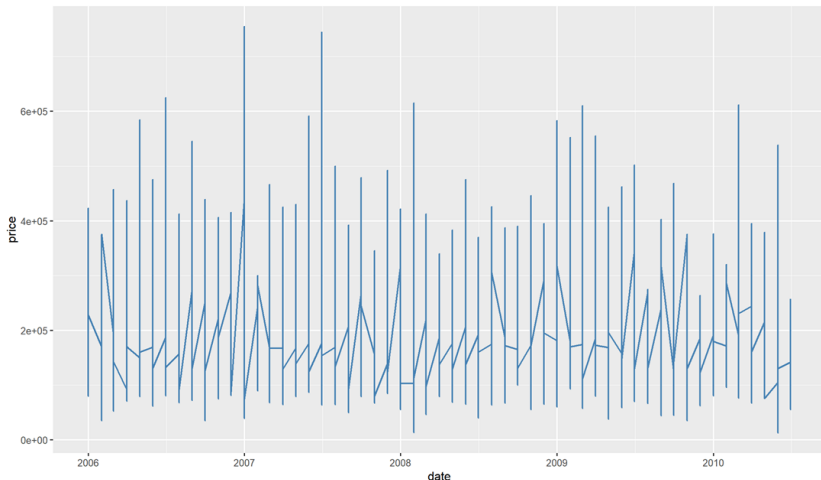- **Example:**

```
# add "01" to date so it is of the form 2025-06-01
housing$date <- as.Date(paste(housing$`Yr.Sold`,
                              housing$`Mo.Sold`,
                              "01",
                              sep = "-"))

ggplot(yearly_birth, aes(x = date, y = price)) +
  geom_line(linewidth = 0.7, color = "steelblue")
```

- Is this enough?

# Draw Line Graph: geom_line()

- What do we miss here?

# Draw Line Graph: geom_line()

- Each vertical line shows the max and min within that date!
    - What if we want to show the mean of the housing price?
    - Use stat = "summary", fun = "mean" in geom_line()
- stat: "summary" means to reduce to one $y$ per $x$.
    - Different type of stat: = "count", = "density", ...
- fun: determines how $y$ will be collapsed into a single point.
    - By default, it uses the "mean". Also have: "median", "sum" ...
    - "fun" has be used together with stat = "summary".
- **Syntax:**

```
ggplot(housing, aes(x = date, y = price)) +
    geom_line(stat = "summary",
              fun = "mean",
              linewidth = 0.7,
              color = "steelblue")
```

# About the R Quiz (Quiz 3)

- Quiz 3 will be held on October 14.
- The R Quiz will cover Week 1-6 materials in the TA Session (supplement included).
- All of the problems will be multiple-choice questions.
    - 15 questions, 2 points each.
- The Quiz is a written exam, so no laptop is needed.
- There will be no TA session in the midterm week. Instead, I will extend the office hours to 16:30 - 19:00 on 10/21.
    - You can check your Quiz 3 scores during the office hours.

# Example Questions: Easy

- (2 points) Which functions should you use if you want to adjust the axes scales for the plots?
  - (a) `rm()`
  - (b) `addmargins()`
  - (c) `labs()`
  - (d) `scale_x_continuous()`

# Example Questions: Normal

- (2 points) Which of the following statements is FALSE?
  - (a) We may use `dotsize = 3` in `geom_point()` to change the size of the points shown in the plot.
  - (b) The output of `seq_along(rnorm(5))` is 1, 2, 3, 4, 5.
  - (c) The output from `replicate()` may not be a matrix.
  - (d) `dnn` is an assignment in `table()` to customize the dimension names.

# Example Questions: Hard

- (2 points) Suppose the output of `as.Date("01-02-0304")` is XXXX-ab-cd. What is ab-cd?
  - (a) 01-02
  - (b) 02-01
  - (c) 02-03
  - (d) 03-04