



# DS 3003: Communicating with Data

Spring 2022

Youmi Suk

School of Data Science, University of Virginia

1.18.2022

# About the Instructor

- Dr. Youmi Suk
- I'm an Assistant Professor in the School of Data Science at UVA trying to develop and apply quantitative models to address practical and important problems in the educational, social, and behavioral sciences. Specifically, my areas of research include causal inference, quasi-experimental designs, multilevel modeling, machine learning methods, and analysis of timing and process data.
- You can find more info about me from <https://youmisuk.com>

# About the TA

- Soumya Chappidi
- Email: [sc4cb@virginia.edu](mailto:sc4cb@virginia.edu)
- Virtual Office Hours:
  - Fridays 9:00 – 10:00 AM

# Introduce Yourself

- Introduce yourself in class.
  - major, year, previous R experiences, and **fun facts!**

# Pre-class survey

- Please complete the **pre-class survey** by **Jan. 23**.

# Syllabus

- Let's review the syllabus.

# Visualizations

- Base R
- ggplot2
- Interactive tools like Plotly and Shiny

# Example data

- mpg data from ggplot2 package
- This dataset contains fuel economy data from 1999 to 2008 for 38 popular models of cars.
- Data frame with 234 rows and 11 variables

```
head(mpg)
```

```
## # A tibble: 6 × 11
##   manufacturer model displ  year   cyl trans      drv    cty   hwy fl
##   <chr>         <chr> <dbl> <int> <int> <chr>    <chr> <int> <int> <chr>
## 1 audi         a4      1.8  1999     4 auto(l5)  f       18    29 p
## 2 audi         a4      1.8  1999     4 manual(m5) f       21    29 p
## 3 audi         a4      2    2008     4 manual(m6) f       20    31 p
## 4 audi         a4      2    2008     4 auto(av)   f       21    30 p
## 5 audi         a4      2.8  1999     6 auto(l5)  f       16    26 p
## 6 audi         a4      2.8  1999     6 manual(m5) f       18    26 p
```

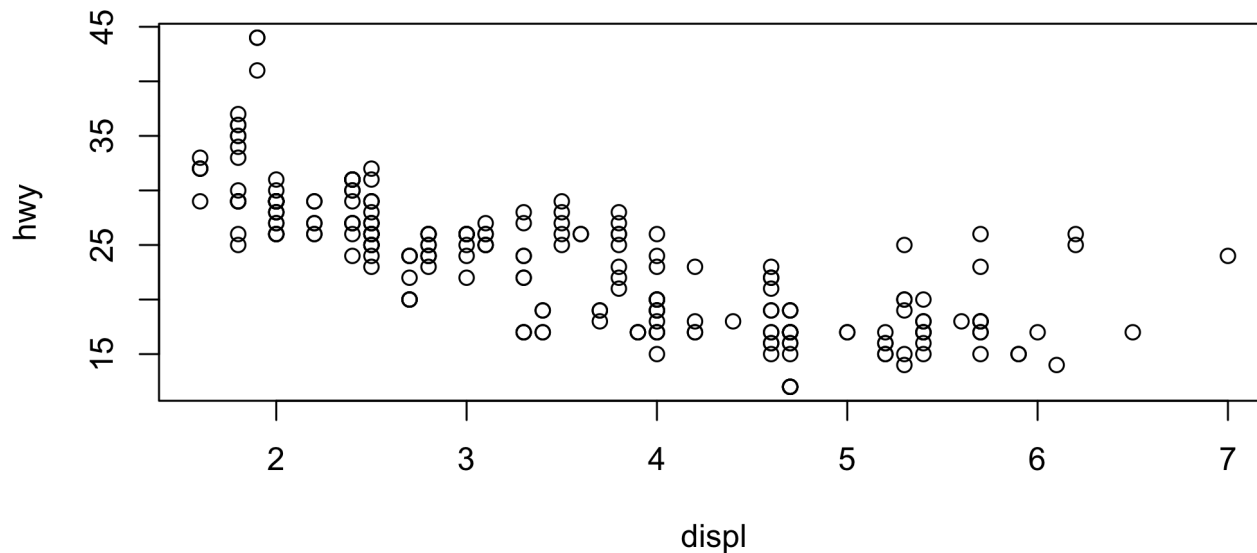
- For more info about mpg data, use ?mpg.



# Base R

- default plotting functions like `plot()` and `hist()`
- scatterplot: *engine displacement by highway miles per gallon*

```
plot(hwy ~ displ, data=mpg)
```

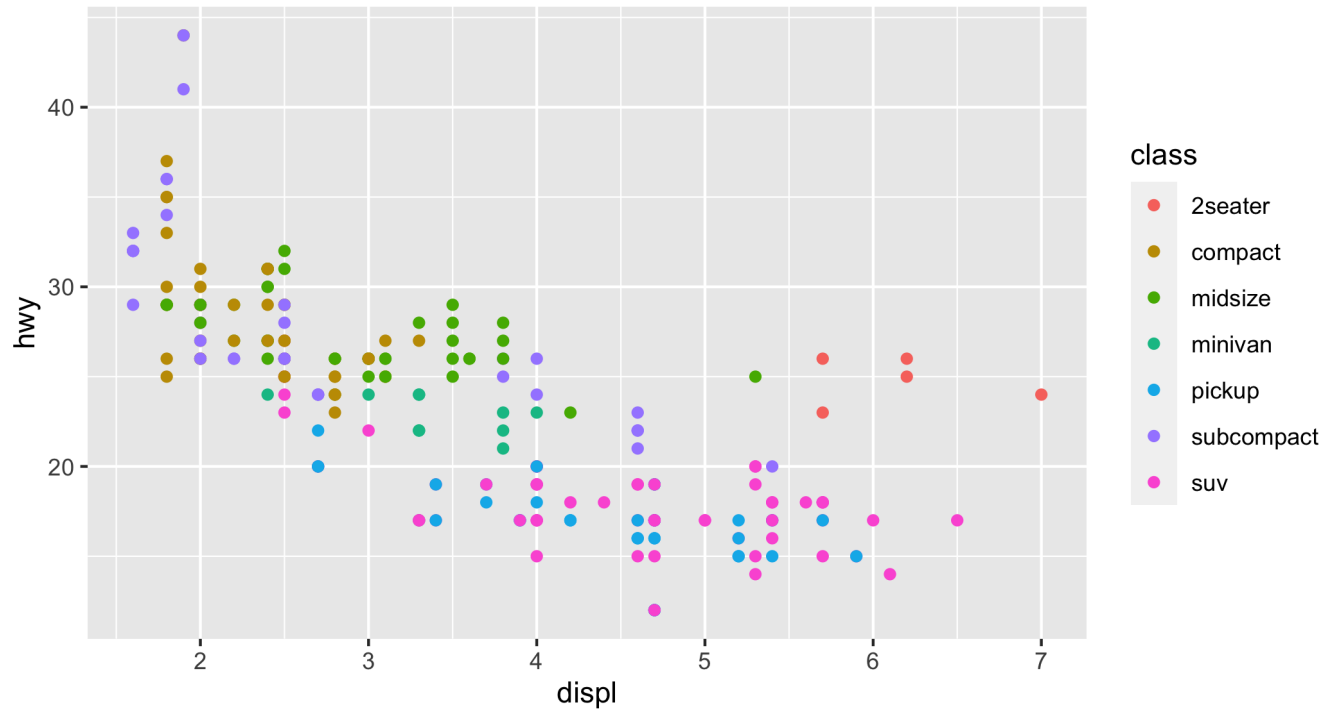


# ggplot2

- ggplot2 is a system for declaratively creating graphics, based on **The Grammar of Graphics**.
- <https://ggplot2.tidyverse.org/index.html>

```
ggplot(mpg, aes(x=displ, y=hwy)) + geom_point()
```

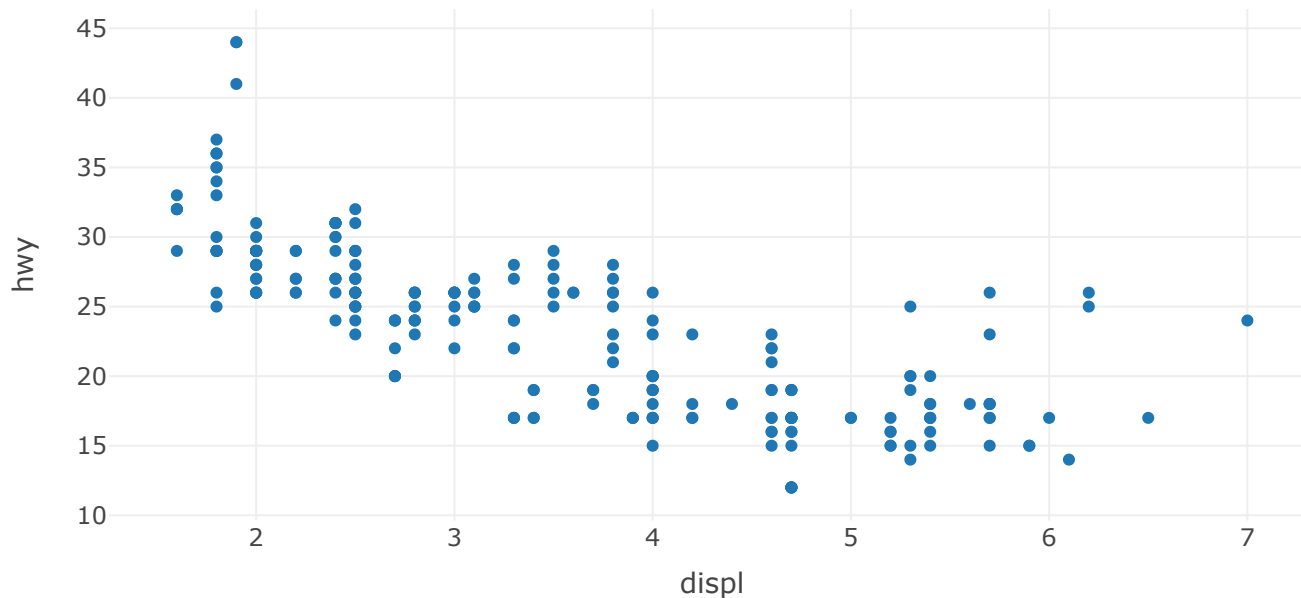
```
ggplot(mpg, aes(x=displ, y=hwy, colour = class)) + geom_point()
```



# plotly

- plotly is an R package for creating interactive web-based graphs.
- <https://plotly.com>

```
fig <- plot_ly(data = mpg, x = ~displ, y = ~hwy, type="scatter", mode  
fig
```



# Shiny

- Shiny is an R package that makes it easy to build interactive web applications (apps) straight from R.
- <https://shiny.rstudio.com>
- An example: <https://youmi.shinyapps.io/curobustml/>

R you ready?