

STATS 306 WI 2020, HW 1, 25 points

Assigned: 01/16/2020

Due: 01/30/2020

Read these instructions completely in order to receive full credit

- Before you submit the problem set, make sure everything runs as expected. Go to the menu bar at the top of Jupyter Notebook and click `Kernel > Restart & Run All`. Your code should run from top to bottom with no errors. Failure to do this will result in loss of points.
- Each problem has their own answer cell. So write or code your answer to their designated cell. If you want to add additional cells make sure to add them just below the designated cell.
- After you done with running your code, use `Ctrl` or `Command` + `p` then click `save as pdf` to save your homework as pdf file. Do not use pdf exporter feature of jupyter! You are not going to upload your `.ipynb` file we just need the `.pdf`.

Your unique id: write here

Problem 1, 5 points

(1) What does **tidy** data mean? 1 point

Type *Markdown* and LaTeX: α^2

(2) **Visualization** and **Modelling** are two main engine of knowledge generation. Write one strength and one weakness of each. 2 points

Type *Markdown* and LaTeX: α^2

(3) **R** is a popular programming language for doing data science. Name two other programming languages often used on data science projects. 1 point

Type *Markdown* and LaTeX: α^2

(4) Name two types of non-rectangular data. *1 point*

Type *Markdown* and LaTeX: α^2

Problem 2, 5 points

(1) Write the command to install the package `tidyverse`. *1 point*

In []:

(2) Write the command to get help on the `mpg` data set. *1 point*

In []:

(3) To create a scatterplot using `displ` and `hwy` variables in the `mpg` data frame, Charlie uses the following command

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

Fix Charlie's code and type it below. *2 points*

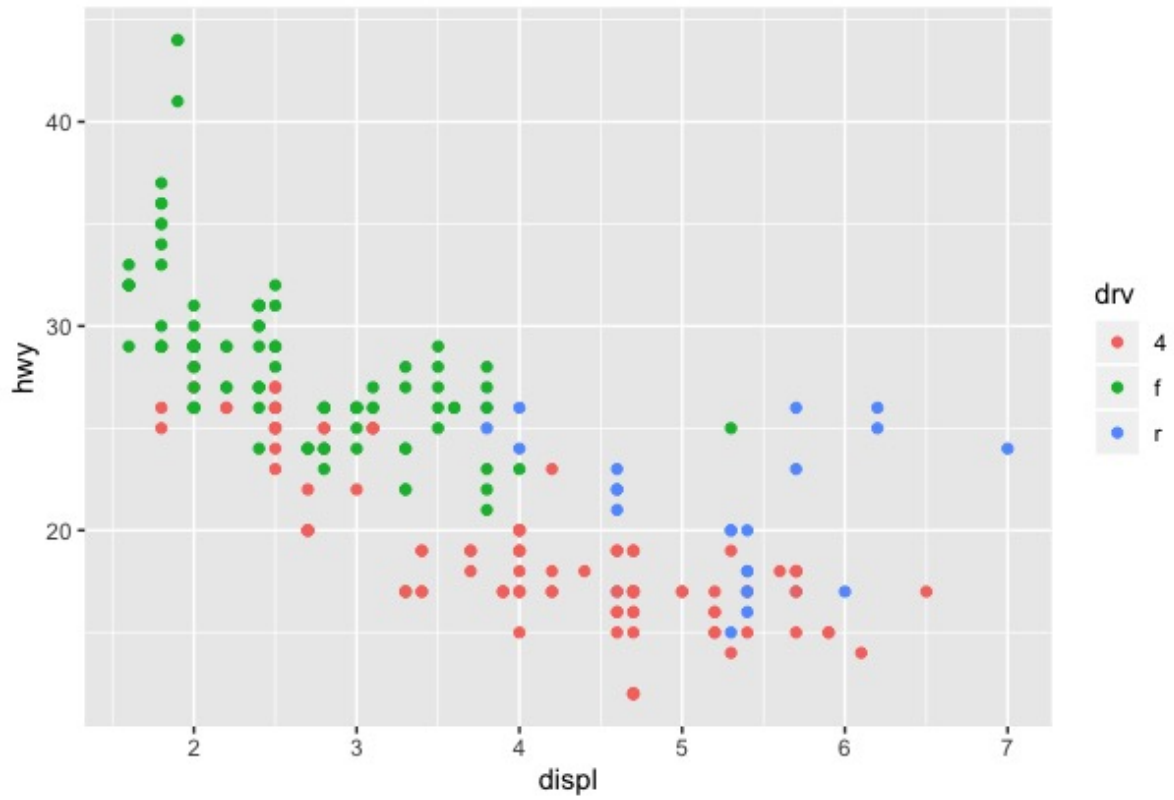
In []:

(4) How many rows and columns does the `mpg` data set have? *1 point*

Type *Markdown* and LaTeX: α^2

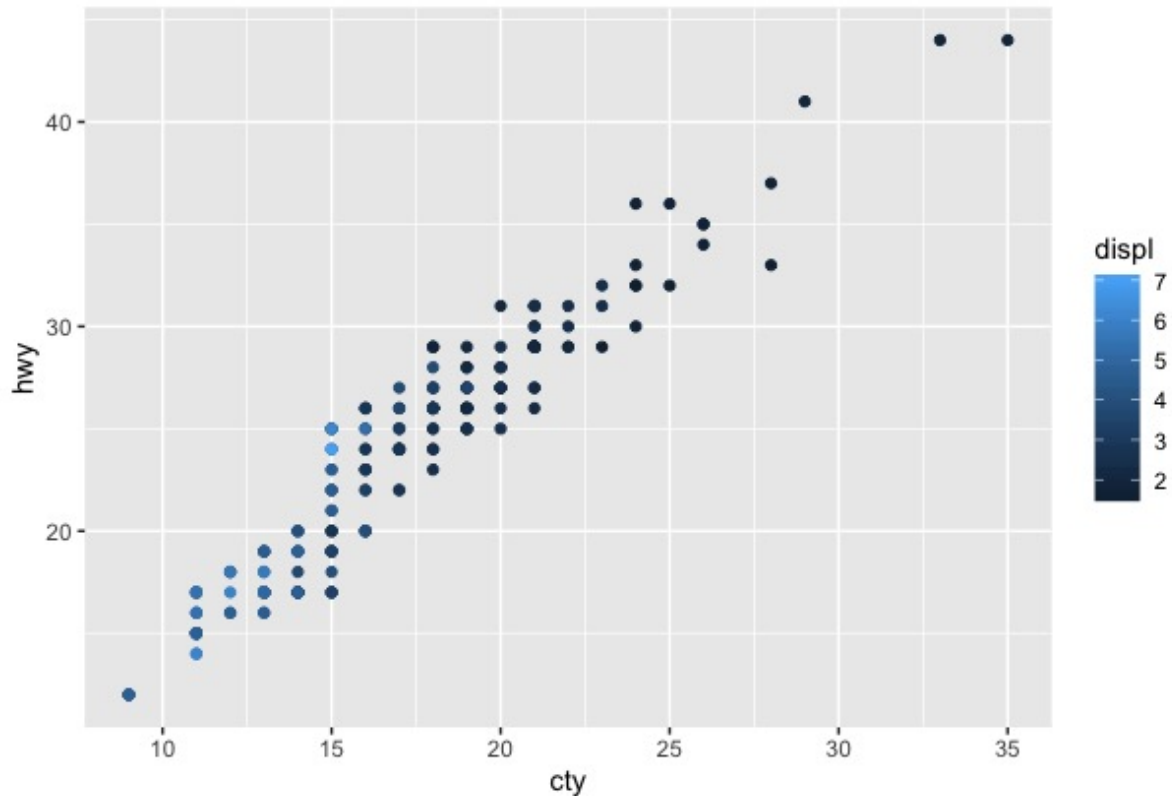
Problem 3, 6 points

(1) Write the code to generate the following plot using the `mpg` data set. *3 points*



In []:

(2) Write the code to generate the following plot using the `mpg` data set. 3 points



In []:

Problem 4, 4 points

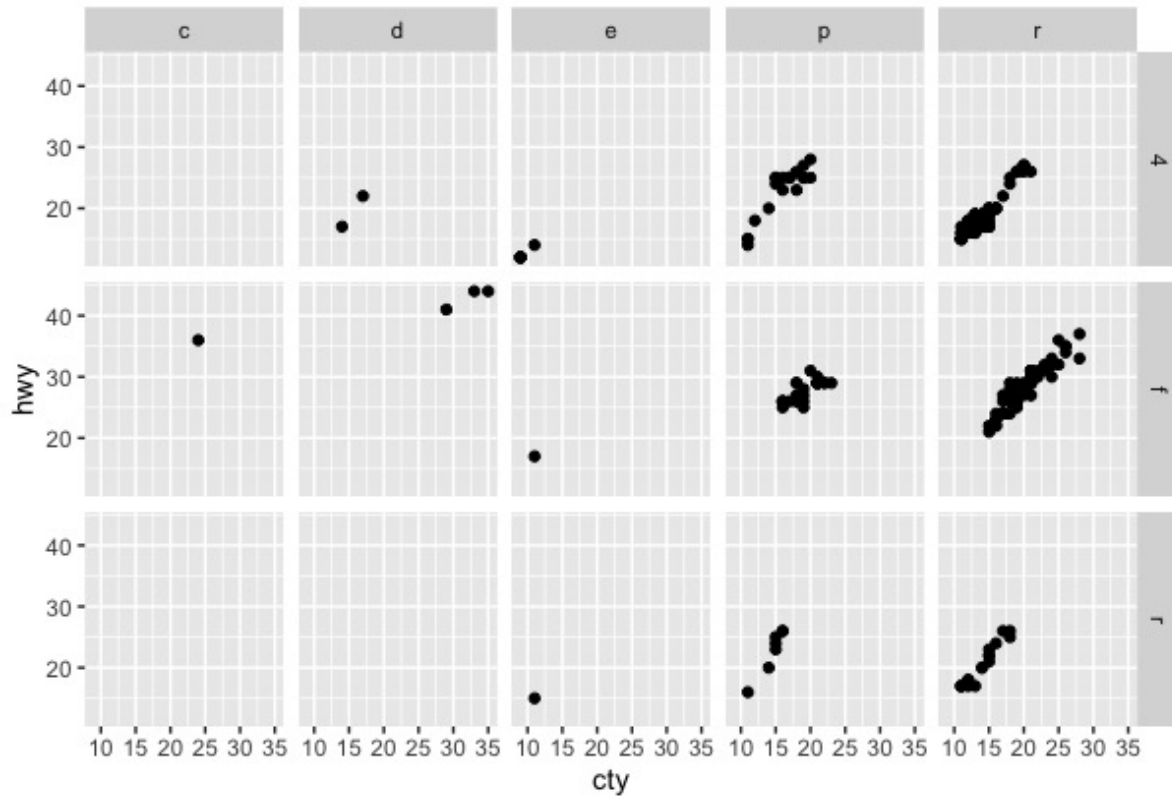
(1) Alice wanted to create a scatter plot using the `mpg` data sets with `displ` on the x-axis and `hwy` on the y-axis. However, she wanted the points to be in blue color. She wrote the following code.

```
ggplot(data = mpg) +  
  geom_point(mapping = aes(x = displ, y = hwy, color = "blue"))
```

This however doesn't quite achieve what Alice was trying to do. Can you help her? Write the correct code below. 2 points

In []:

(2) Write the code to generate the following plot using the `mpg` data set. 2 points



In []:

Problem 5, 5 points

(1) What does the `se` argument to `geom_smooth()` do? 1 point

Type *Markdown* and LaTeX: α^2

(2) Write the code to generate the following plot using the `mpg` data set. 4 points

