

BIOS512 Assignment #1

This assignment will be submitted via GitHub. Once complete, download your notebook and upload it to your GitHub repository for the class.

Note:

- To download a notebook with Jupyter Lab, right click on the notebook in the sidebar file browser on the left and select **Download**.
- To download a notebook in the Jupyter Notebook interface, click **File > Download As > Notebook (.ipynb)**.

Import the `tidyverse` into your R session.

```
In [1]: library(tidyverse)
```

— **Attaching packages** — tidyverse 1.2.1 —

```
✓ ggplot2 3.2.0      ✓ purrr 0.3.2
✓ tibble 2.1.3       ✓ dplyr 0.8.3
✓ tidyr 0.8.3        ✓ stringr 1.4.0
✓ readr 1.3.1        ✓ forcats 0.4.0
```

— **Conflicts** — tidyverse_conflicts() —

```
✗ dplyr::filter() masks stats::filter()
✗ dplyr::lag()     masks stats::lag()
```

What `tidyverse` functions have conflicts with other functions in your R session? (Answer in a markdown cell using the bulleted list syntax)

- `filter()`
- `lag()`

Preview the `mpg` dataset.

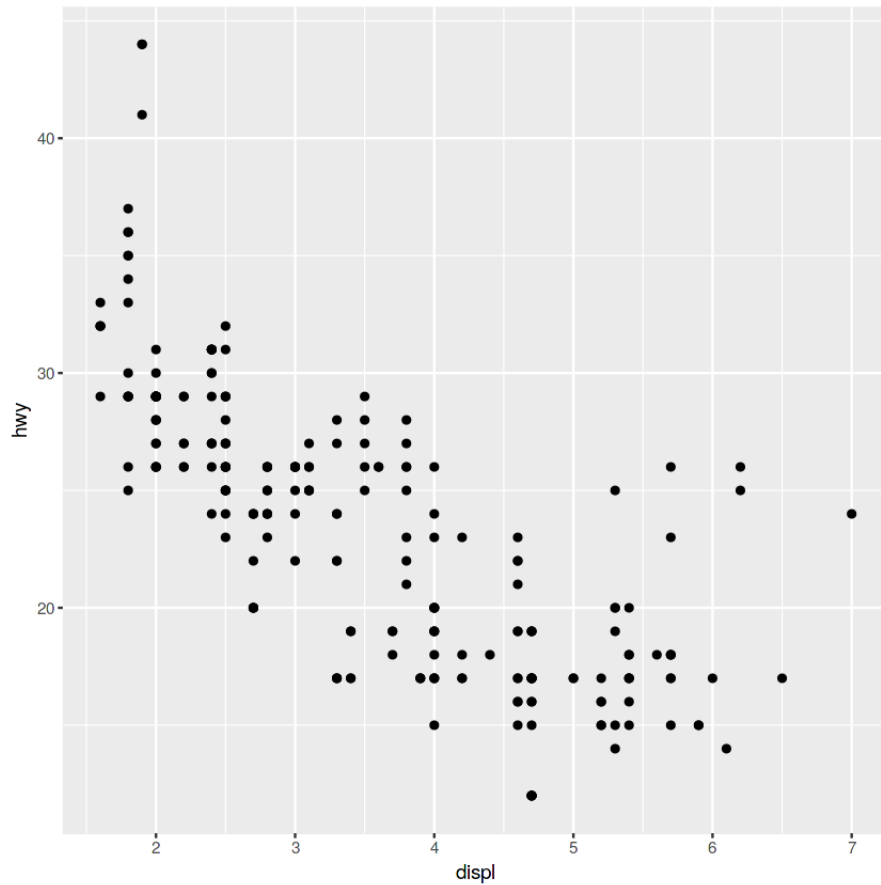
```
In [2]: mpg %>% head
```

A tibble: 6 × 11

manufacturer	model	displ	year	cyl	trans	drv	cty	hwy	fl	class
<chr>	<chr>	<dbl>	<int>	<int>	<chr>	<chr>	<int>	<int>	<chr>	<chr>
audi	a4	1.8	1999	4	auto(l5)	f	18	29	p	compact
audi	a4	1.8	1999	4	manual(m5)	f	21	29	p	compact
audi	a4	2.0	2008	4	manual(m6)	f	20	31	p	compact
audi	a4	2.0	2008	4	auto(av)	f	21	30	p	compact
audi	a4	2.8	1999	6	auto(l5)	f	16	26	p	compact
audi	a4	2.8	1999	6	manual(m5)	f	18	26	p	compact

Using the `mpg` dataset, make a scatter chart with `displ` on the x-axis and `hwy` on the y-axis.

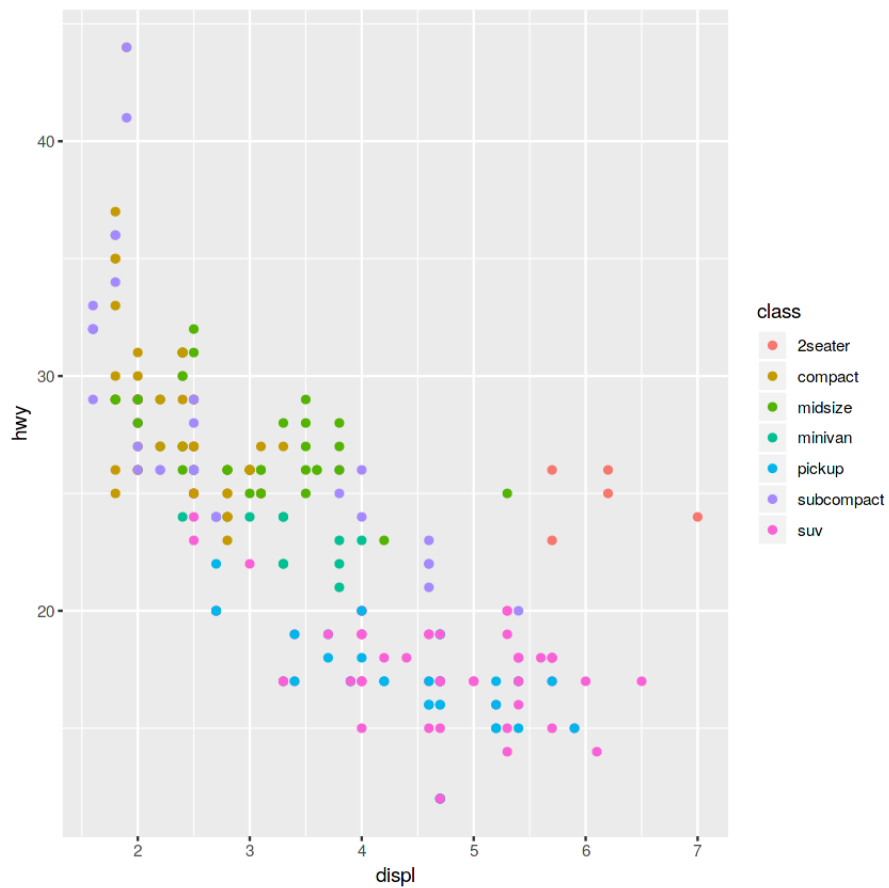
```
In [4]: p = ggplot(mpg, aes(x=displ, y=hwy)) +  
        geom_point()  
  
p
```



Color the points from your plot by `class` .

```
In [5]: p = ggplot(mpg, aes(x=displ, y=hwy, color=class)) +  
        geom_point()
```

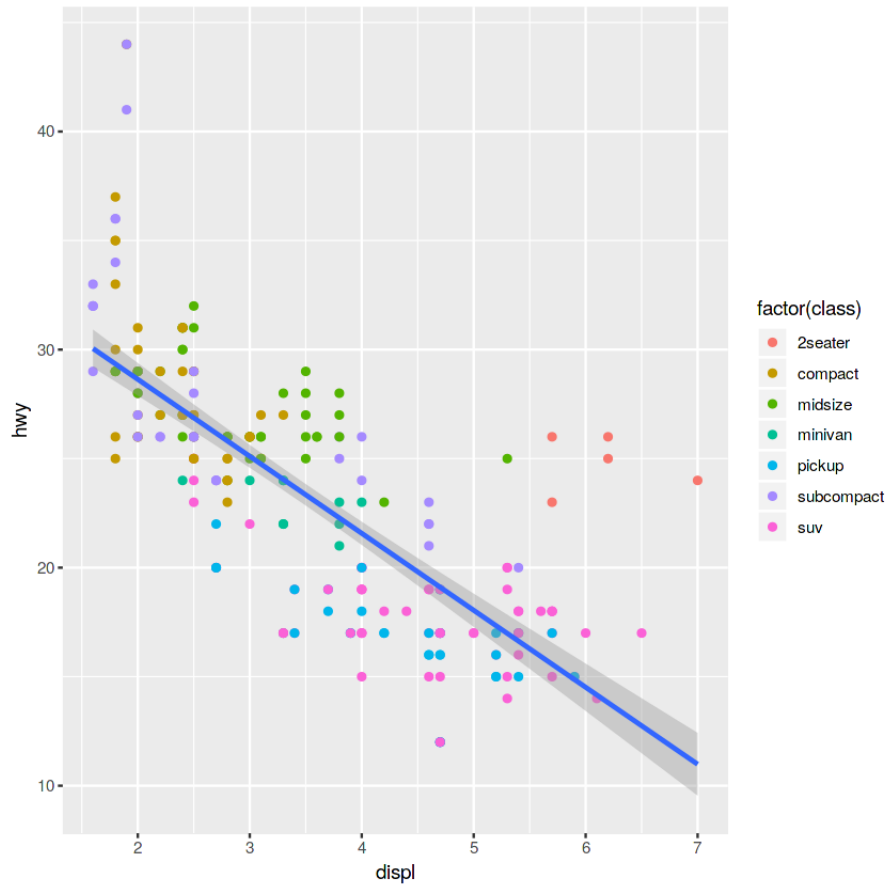
p



Add a smooth geom to your mpg plot.

```
In [11]: p = ggplot(mpg, aes(x=displ, y=hwy)) +  
  geom_point(aes(color=factor(class))) +  
  geom_smooth(method = 'lm')
```

p



What are two chart features you can adjust in `ggplot2` to help with overplotting? (Answer in a markdown cell using the **numbered** list syntax)

- `alpha` (changes transparency of points)
- `jitter` (changes position of points)