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 Jupter Notebook interface, click `File > Download As > Notebook (.ipynb)`.

Import the `tidyverse` into your R session.

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

```
── Attaching packages ─────────────────────────────────────── tidyvers
e 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_conf
licts() ──
✖ 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 list syntax)

- filter()
- lag()

Preview the `mpg` dataset.

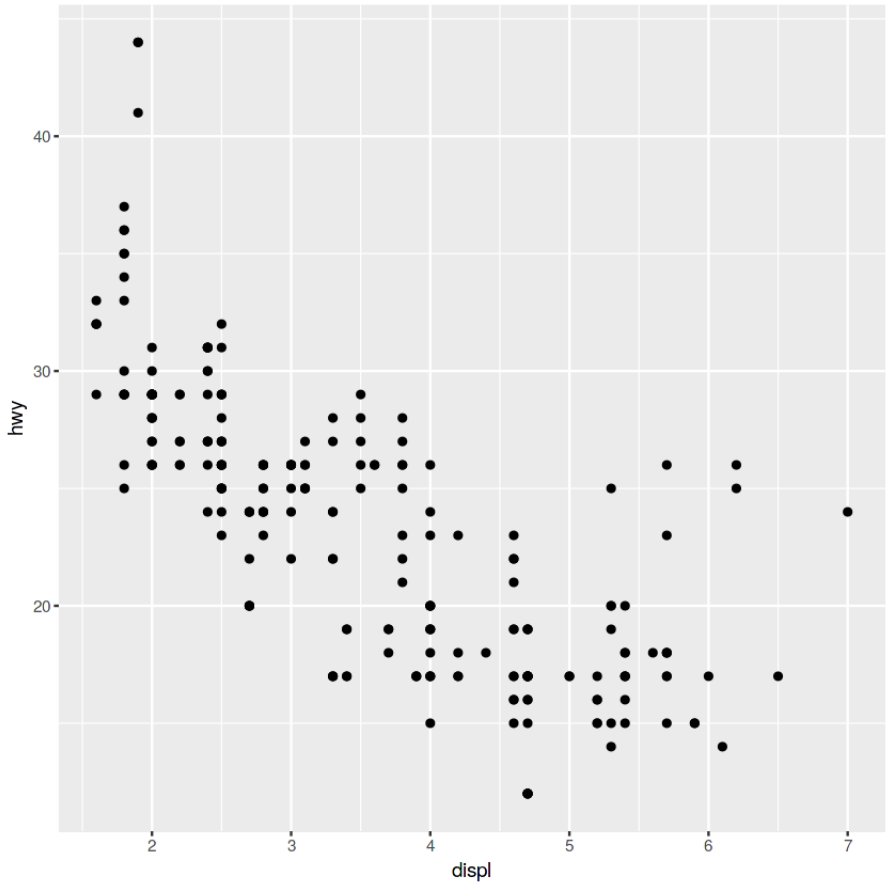
In [2]:
```
mpg %>% head
```

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

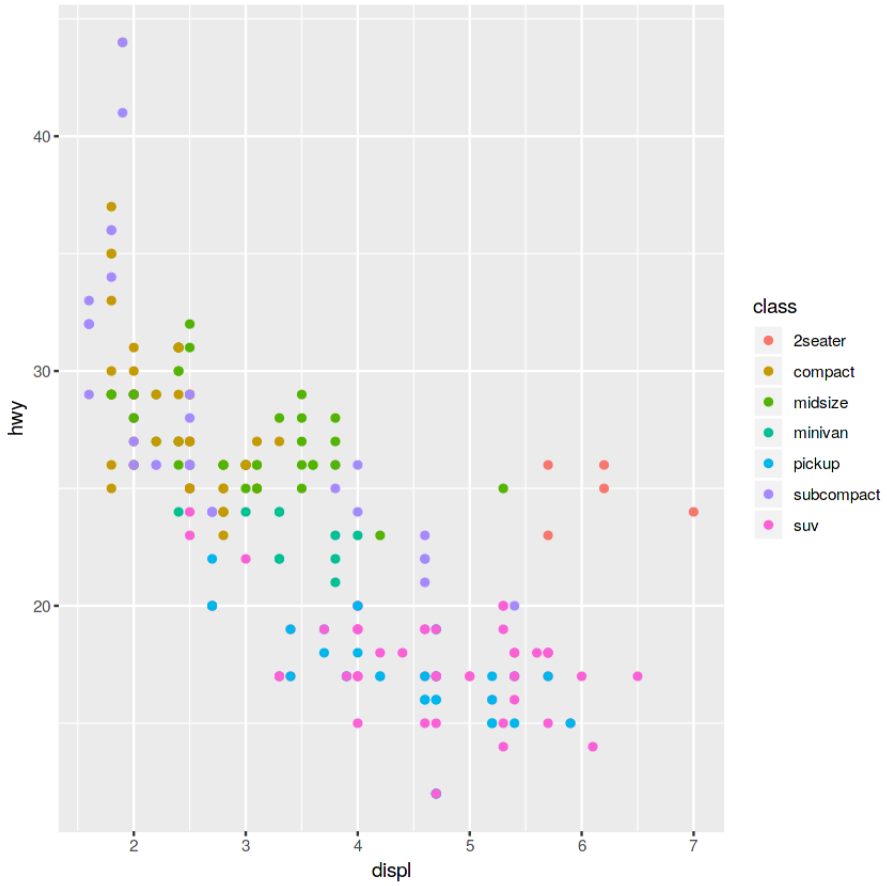
Using the `mpg` dataset, make a scatter chart with `displ` on the x-axis and `hwy` on the y-axis.

In [3]:
```
p = ggplot(mpg, aes(x=displ, y=hwy)) +
    geom_point()

p
```



Color the points from your plot by `class`.

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

p
```



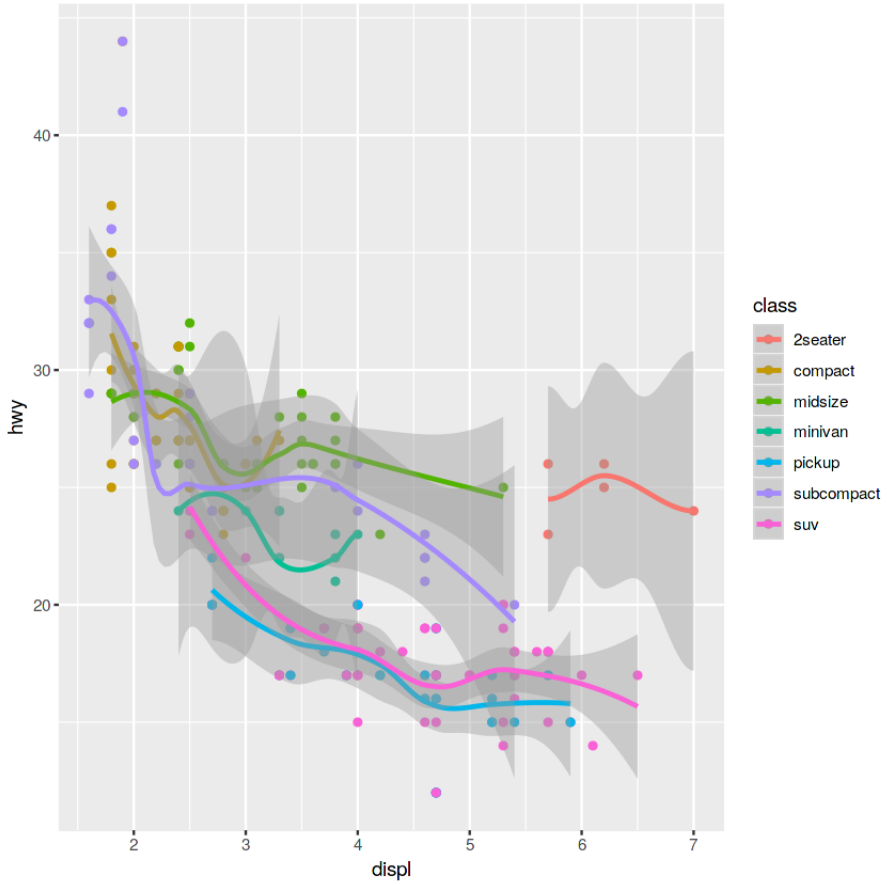
Add a `smooth` geom to your `mpg` plot.

In [9]:
```
p = p + geom_smooth()

p
```

```
`geom_smooth()` using method = 'loess' and formula 'y ~ x'

Warning message in simpleLoess(y, x, w, span, degree = degree, paramet
ric = parametric, :
“span too small.   fewer data values than degrees of freedom.”
Warning message in simpleLoess(y, x, w, span, degree = degree, paramet
ric = parametric, :
“pseudoinverse used at 5.6935”
Warning message in simpleLoess(y, x, w, span, degree = degree, paramet
ric = parametric, :
“neighborhood radius 0.5065”
Warning message in simpleLoess(y, x, w, span, degree = degree, paramet
```

```
ric = parametric, :
“reciprocal condition number  0”
Warning message in simpleLoess(y, x, w, span, degree = degree, paramet
ric = parametric, :
“There are other near singularities as well. 0.65044”
Warning message in predLoess(object$y, object$x, newx = if (is.null(ne
wdata)) object$x else if (is.data.frame(newdata)) as.matrix(model.fram
e(delete.response(terms(object)), :
“span too small.   fewer data values than degrees of freedom.”
Warning message in predLoess(object$y, object$x, newx = if (is.null(ne
wdata)) object$x else if (is.data.frame(newdata)) as.matrix(model.fram
e(delete.response(terms(object)), :
“pseudoinverse used at 5.6935”
Warning message in predLoess(object$y, object$x, newx = if (is.null(ne
wdata)) object$x else if (is.data.frame(newdata)) as.matrix(model.fram
e(delete.response(terms(object)), :
“neighborhood radius 0.5065”
Warning message in predLoess(object$y, object$x, newx = if (is.null(ne
wdata)) object$x else if (is.data.frame(newdata)) as.matrix(model.fram
e(delete.response(terms(object)), :
“reciprocal condition number  0”
Warning message in predLoess(object$y, object$x, newx = if (is.null(ne
wdata)) object$x else if (is.data.frame(newdata)) as.matrix(model.fram
e(delete.response(terms(object)), :
“There are other near singularities as well. 0.65044”
Warning message in simpleLoess(y, x, w, span, degree = degree, paramet
ric = parametric, :
“pseudoinverse used at 4.008”
Warning message in simpleLoess(y, x, w, span, degree = degree, paramet
ric = parametric, :
“neighborhood radius 0.708”
Warning message in simpleLoess(y, x, w, span, degree = degree, paramet
ric = parametric, :
“reciprocal condition number  0”
Warning message in simpleLoess(y, x, w, span, degree = degree, paramet
ric = parametric, :
“There are other near singularities as well. 0.25”
Warning message in predLoess(object$y, object$x, newx = if (is.null(ne
wdata)) object$x else if (is.data.frame(newdata)) as.matrix(model.fram
e(delete.response(terms(object)), :
“pseudoinverse used at 4.008”
Warning message in predLoess(object$y, object$x, newx = if (is.null(ne
wdata)) object$x else if (is.data.frame(newdata)) as.matrix(model.fram
e(delete.response(terms(object)), :
“neighborhood radius 0.708”
Warning message in predLoess(object$y, object$x, newx = if (is.null(ne
wdata)) object$x else if (is.data.frame(newdata)) as.matrix(model.fram
e(delete.response(terms(object)), :
“reciprocal condition number  0”
Warning message in predLoess(object$y, object$x, newx = if (is.null(ne
```

```
wdata)) object$x else if (is.data.frame(newdata)) as.matrix(model.fram
e(delete.response(terms(object)), :
“There are other near singularities as well. 0.25”
```



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

1. position = "jitter", to adjust position of the points
2. alpha = 0.2, to increase transparency of the points