

Class 05: Data Visualization with GGLOT

Mady Welch

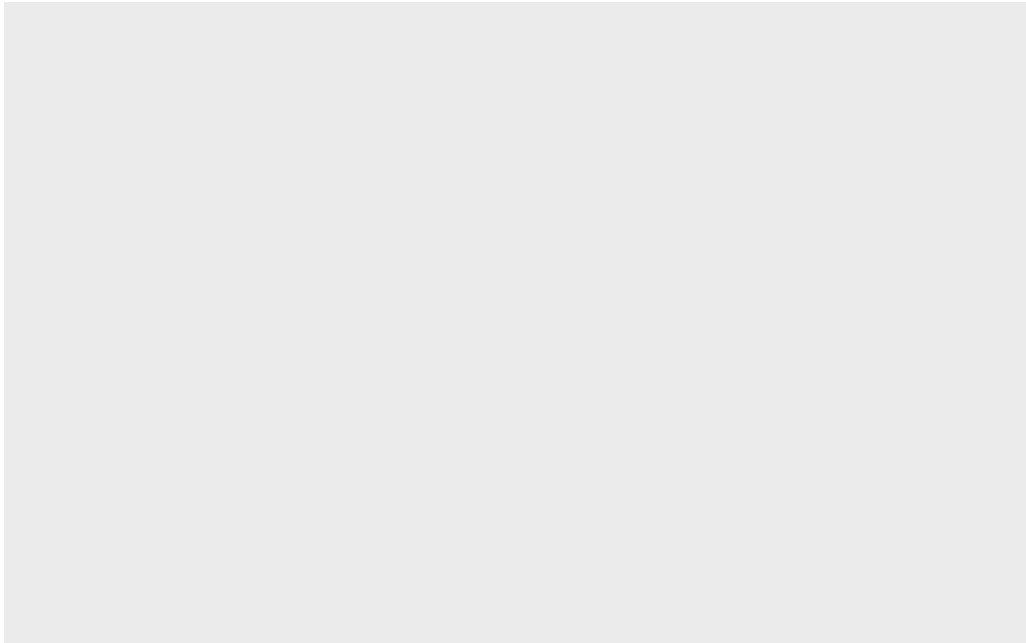
Our First ggplot

To use the ggplot2 package first need to have it installed on my computer.

To install package we use the `install.packages()` command.

Can't use it until we use `library(ggplot2)`

```
library(ggplot2)  
ggplot()
```



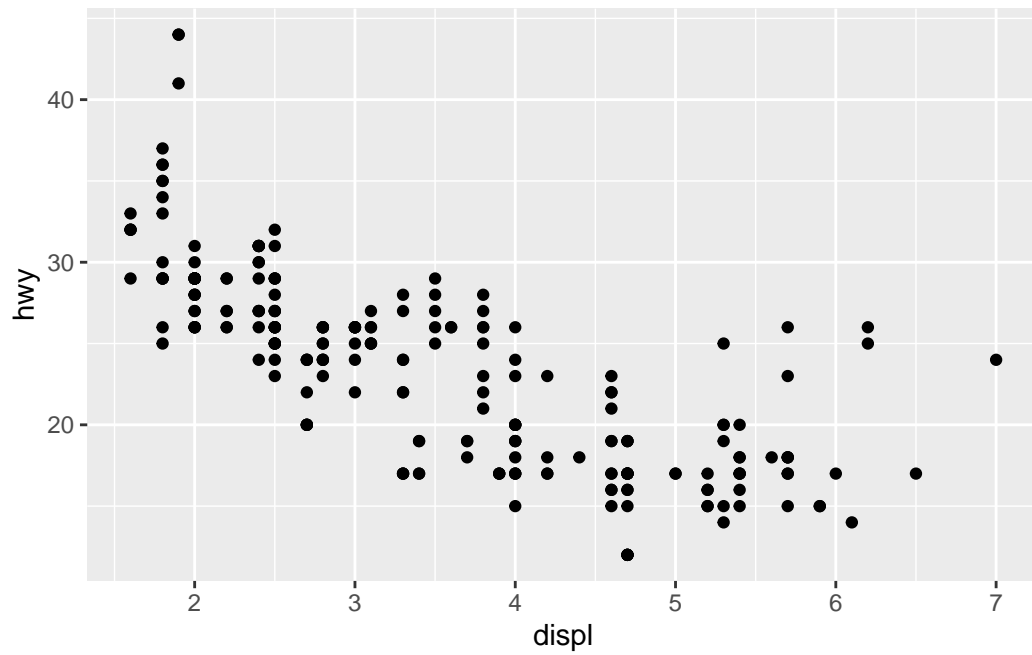
```
mpg
```

```
# A tibble: 234 x 11
  manufacturer model      displ  year   cyl trans drv     cty   hwy fl      class
  <chr>          <chr>    <dbl> <int> <int> <chr> <chr> <int> <int> <chr> <chr>
1 audi          a4         1.8  1999     4 auto~ f      18    29 p    comp~
2 audi          a4         1.8  1999     4 manu~ f      21    29 p    comp~
3 audi          a4         2    2008     4 manu~ f      20    31 p    comp~
4 audi          a4         2    2008     4 auto~ f      21    30 p    comp~
5 audi          a4         2.8  1999     6 auto~ f      16    26 p    comp~
6 audi          a4         2.8  1999     6 manu~ f      18    26 p    comp~
7 audi          a4         3.1  2008     6 auto~ f      18    27 p    comp~
8 audi          a4 quattro 1.8  1999     4 manu~ 4      18    26 p    comp~
9 audi          a4 quattro 1.8  1999     4 auto~ 4      16    25 p    comp~
10 audi         a4 quattro  2    2008     4 manu~ 4      20    28 p    comp~
# ... with 224 more rows
```

Our first plot of displ vs. hwy All ggplot() graphs are made in the same way.

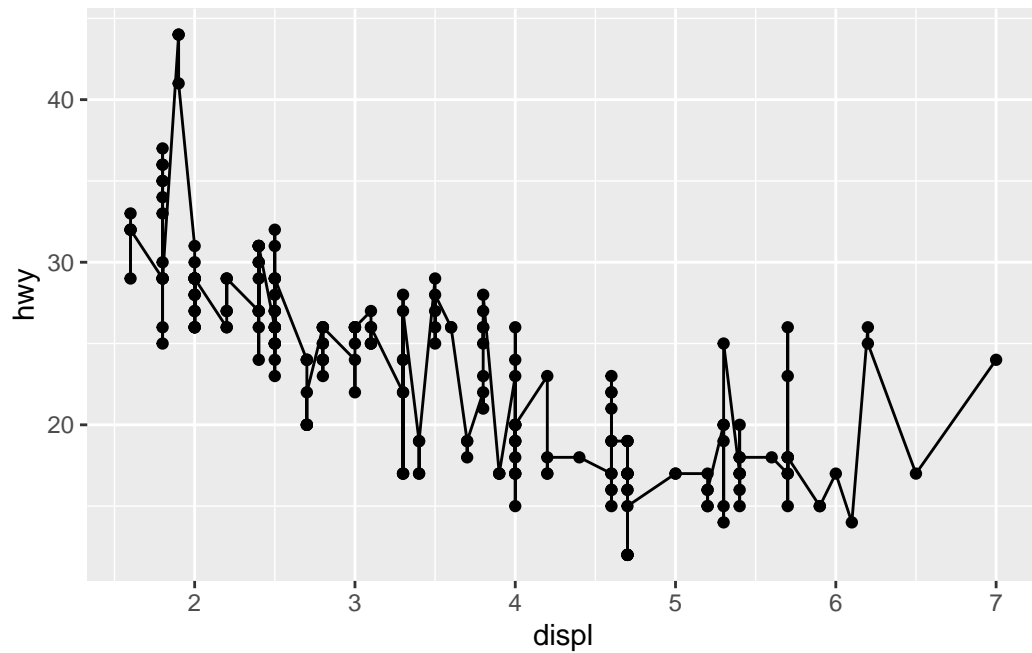
- data + aes + geoms

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



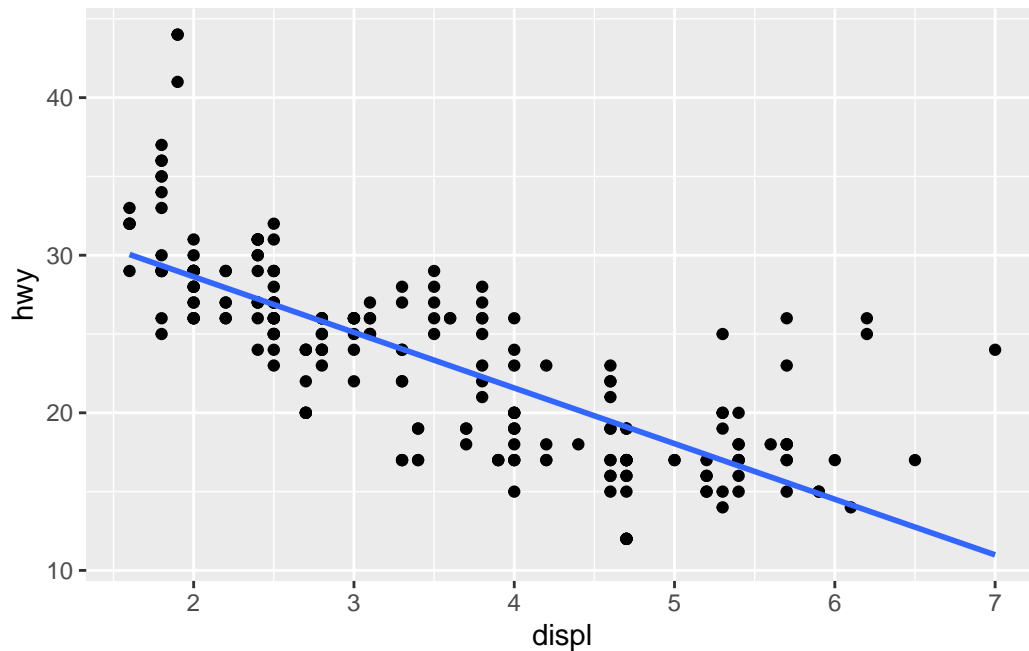
Can add more layers:

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



```
ggplot(mpg) +  
  aes(x=displ, y=hwy) +  
  geom_point() +  
  geom_smooth(method = lm, se = FALSE)
```

`geom_smooth()` using formula 'y ~ x'



Plot of Gene Expression Data

First read data from online and store it in environment.

```
url <- "https://bioboot.github.io/bimm143_S20/class-material/up_down_expression.txt"
genes <- read.delim(url)
head(genes)
```

	Gene	Condition1	Condition2	State
1	A4GNT	-3.6808610	-3.4401355	unchanging
2	AAAS	4.5479580	4.3864126	unchanging
3	AASDH	3.7190695	3.4787276	unchanging
4	AATF	5.0784720	5.0151916	unchanging
5	AATK	0.4711421	0.5598642	unchanging
6	AB015752.4	-3.6808610	-3.5921390	unchanging

Q. How many genes?

```
nrow(genes)
```

```
[1] 5196
```

Q. What are the columns?

```
colnames(genes)
```

```
[1] "Gene"          "Condition1" "Condition2" "State"
```

Q. How many upregulated genes? What fraction of total genes is upregulated in this dataset (2 sigfigs)?

```
table(genes[, "State"])
```

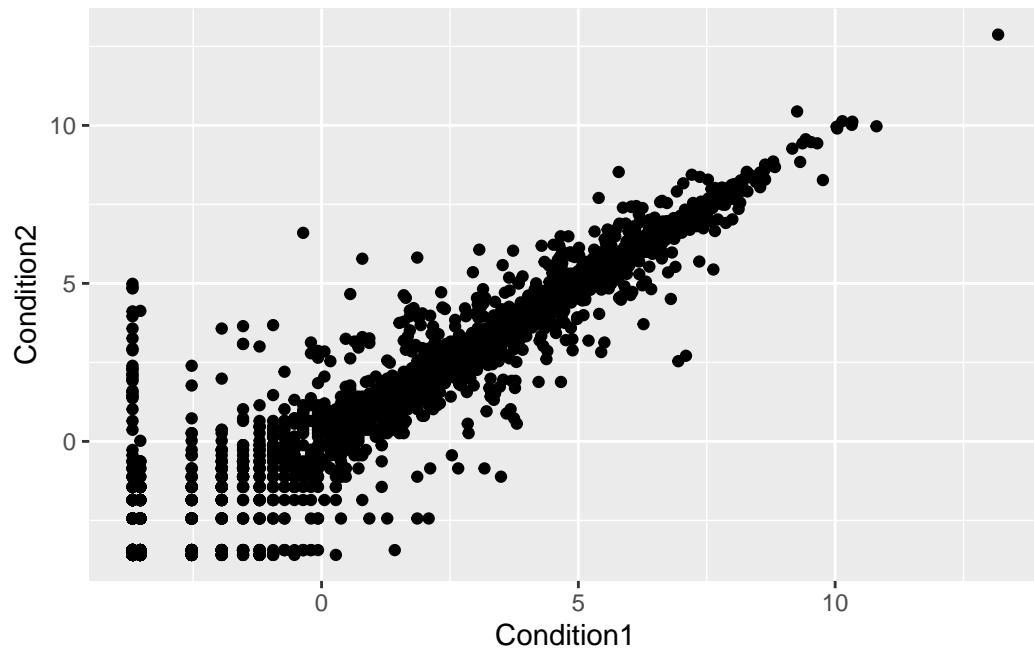
down	unchanging	up
72	4997	127

```
round(table(genes[, "State"])/nrow(genes)*100, 2)
```

down	unchanging	up
1.39	96.17	2.44

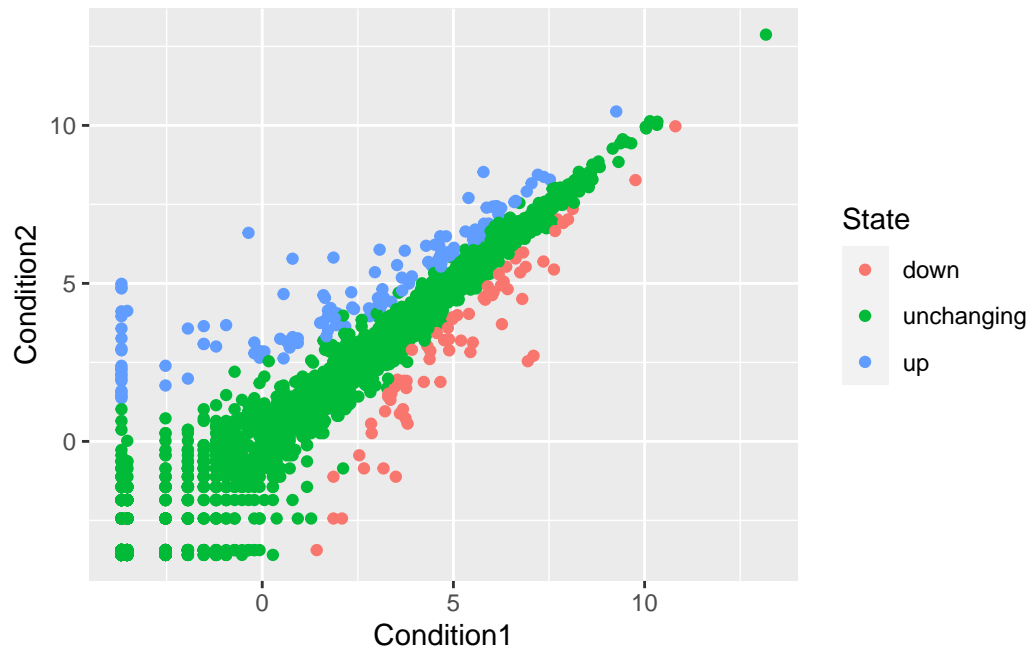
A first version plot of this data Condition1 vs Condition2

```
ggplot(genes) +  
  aes(x=Condition1, y=Condition2) +  
  geom_point()
```



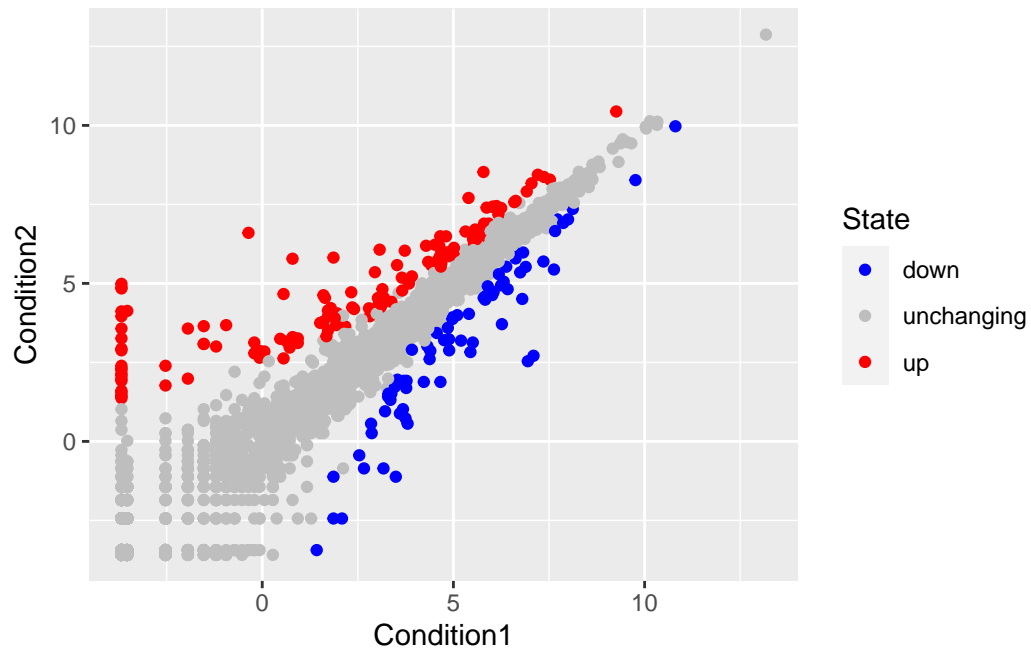
Add Color:

```
p <- ggplot(genes) +  
  aes(x=Condition1, y=Condition2, col=State) +  
  geom_point()  
p
```



Change colors:

```
p + scale_colour_manual(values=c("blue", "gray", "red"))
```

Add plot annotations using `labs()` function:

```
p + scale_colour_manual(values=c("blue", "gray", "red")) +  
  labs(title="Gene Expression Changes Upon Drug Treatment", x="Control (no drug)", y="Drug")
```

Gene Expression Changes Upon Drug Treatment

