Class 5: Data Visualization

Using the ggplot2 package

Meg Robinson

2022-02-05

```r
####################################
# Author: Meg Robinson
# Class 5
# Section 5: Scatter Plots
####################################


# if not already, install ggplot2

# install.packages("ggplot2")


# load the ggplot2 package after installation
library(ggplot2)

# Use the built-in 'cars' dataset for visualization

# Look at whats in the dataset
head(cars)

##    speed dist
## 1      4    2
## 2      4   10
## 3      7    4
## 4      7   22
## 5      8   16
## 6      9   10

# Use ggplots 3 layers: data, aes, geoms
# To visualize 'cars' as a scatterplot
# Use + to add layers
plot = ggplot(cars) +
  aes(x=speed, y=dist) + geom_point()
plot
```
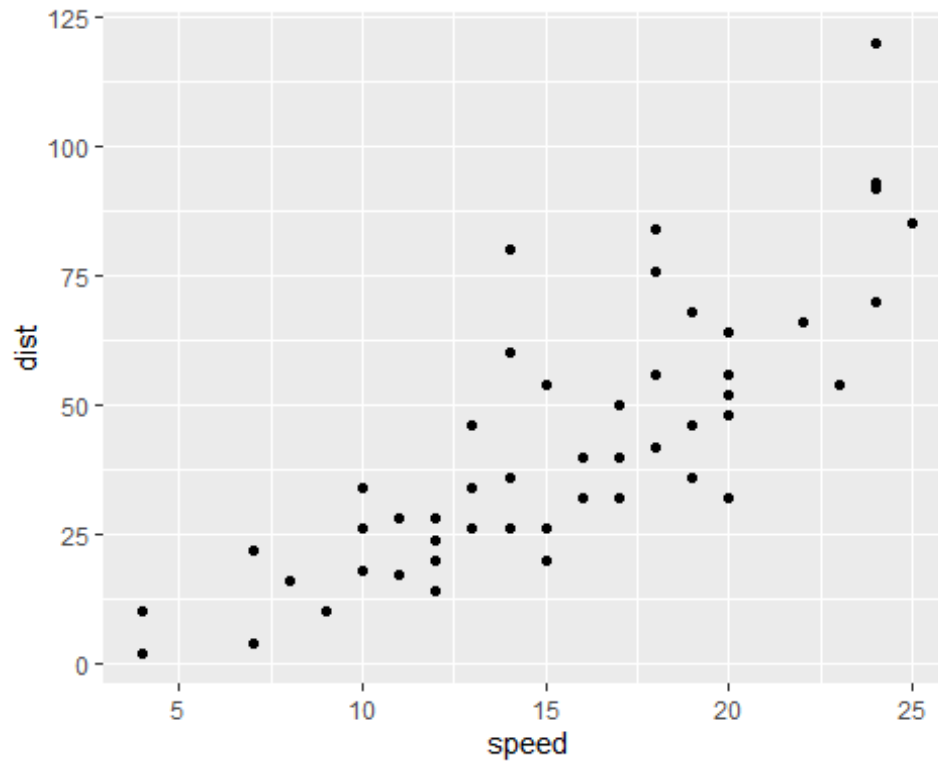
```
# Add geom_smooth() to show the relationship between variables
plot = plot + geom_smooth(method='lm', se=FALSE)
plot

## `geom_smooth()` using formula 'y ~ x'
```
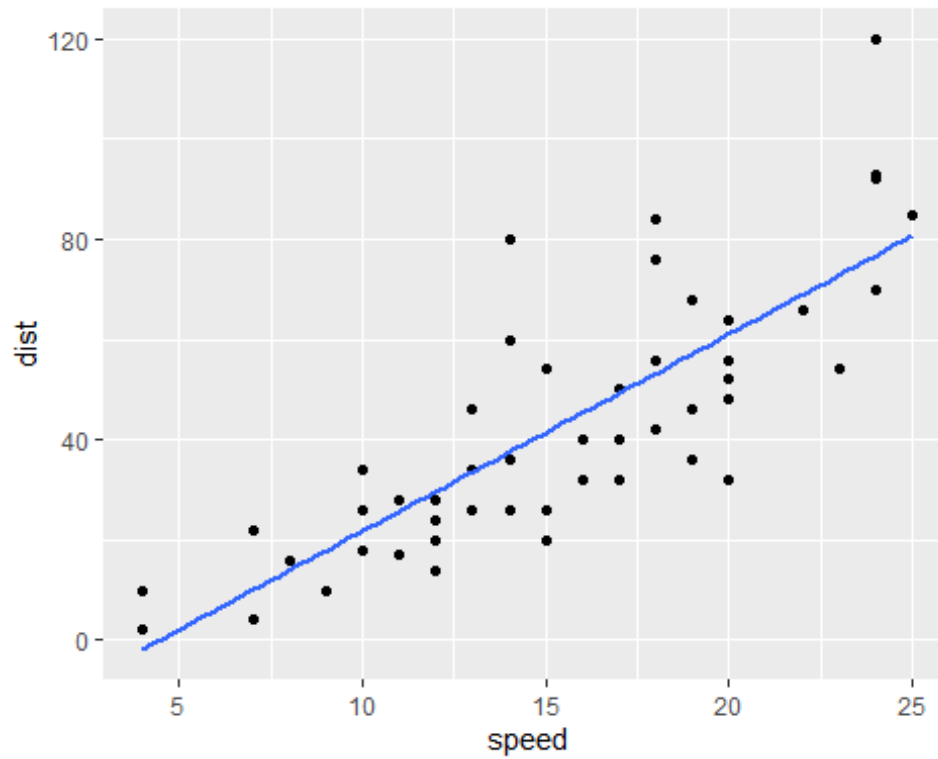
```
plot = plot + labs(title='Speed vs Distance of Cars',
                   x="Speed (mph)",
                   y="Distance to stop (ft)")  +
  theme_bw()
plot

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

Speed vs Distance of Cars
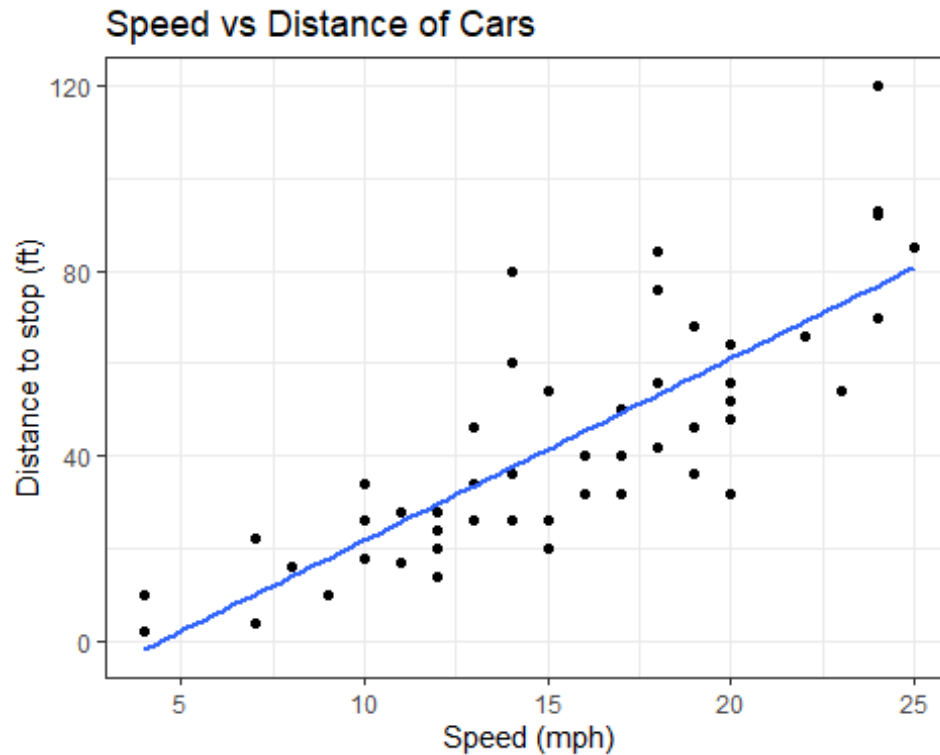

```r
# Read in the results of DE analysis
url <- "https://bioboot.github.io/bimm143_S20/class-
material/up_down_expression.txt"
genes <- read.delim(url)
head(genes)
```

```
##          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
```

```r
# How many genes are in the dataset?
nrow(genes)
```

```
## [1] 5196
```

```r
# How many columns are in the dataset?
ncol(genes)
```

```
## [1] 4
```

```r
# What are the column names?
colnames(genes)
```

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

```
# How many upregulated genes are there?
table(genes$State)

##
##      down unchanging        up
##        72       4997       127

# What percent of up regulated genes?
round(table(genes$State)/nrow(genes)*100,2)

##
##      down unchanging        up
##      1.39      96.17      2.44

# Make a scatter plot of the genes df
ggplot(genes) +
  aes(x=Condition1, y=Condition2) +
  geom_point()
```
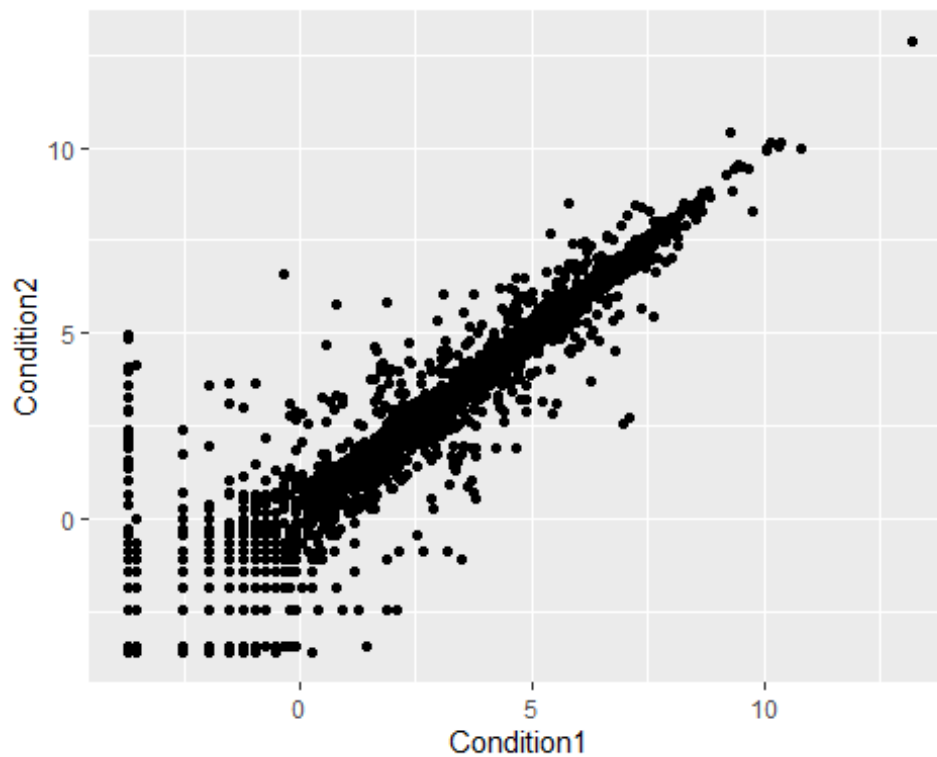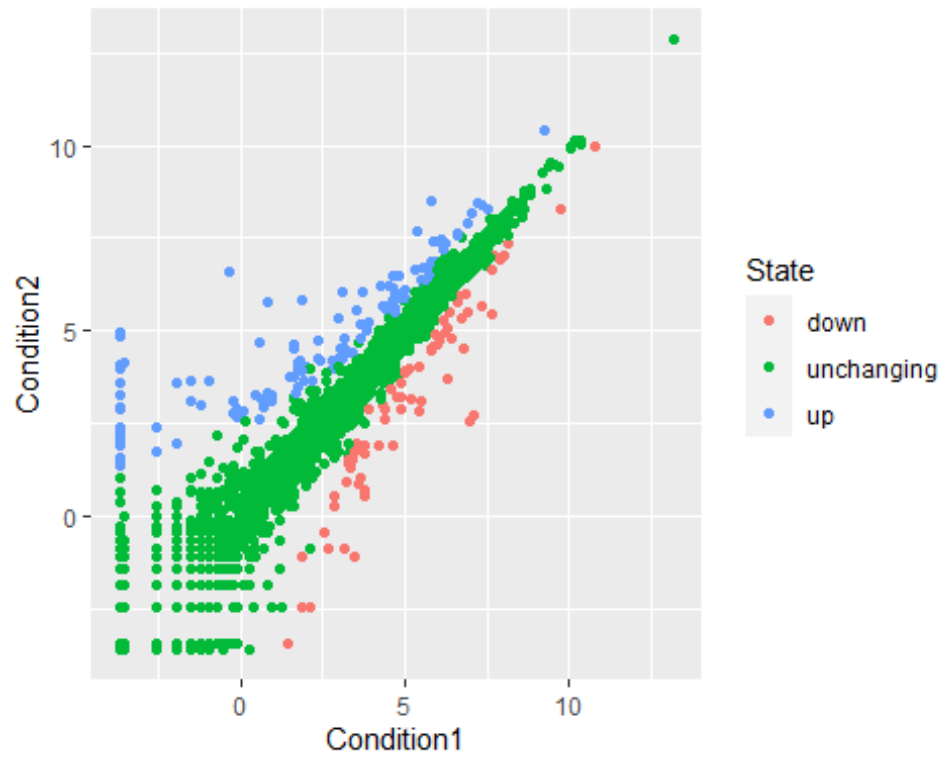


```
# Map the the column  State to color
gene_plot = ggplot(genes) +
  aes(x=Condition1, y=Condition2, col=State) +
  geom_point()
gene_plot
```

```
# add titles etc
gene_plot = gene_plot + theme_bw() +
scale_colour_manual(values=c("blue","gray","red")) +
  labs(title="Gene Expresion Changes Upon Drug Treatment",
       x="Control (no drug) ",
       y="Drug Treatment")
gene_plot
```

Gene Expresion Changes Upon Drug Treatment