

Lab 5–Data Visualization with ggplot

Ramola Baviskar (A12228297)

Feb. 7, 2022

```
#Week 5--Data Visualization with ggplot
#Lab Session

#Q1. For which phases is data visualization important in our scientific workflows?
  #All of the above
#Q2. True or False? The ggplot2 package comes already installed with R?
  #FALSE
#Q3. Which plot types are typically NOT used to compare distributions of numeric variables?
  #Network graph
#Q4. Which statement about data visualization with ggplot2 is incorrect?
  #ggplot is the only way to create plots in R
#Q5. Which geometric layer should be used to create scatter plots in ggplot2?
  #geom_point

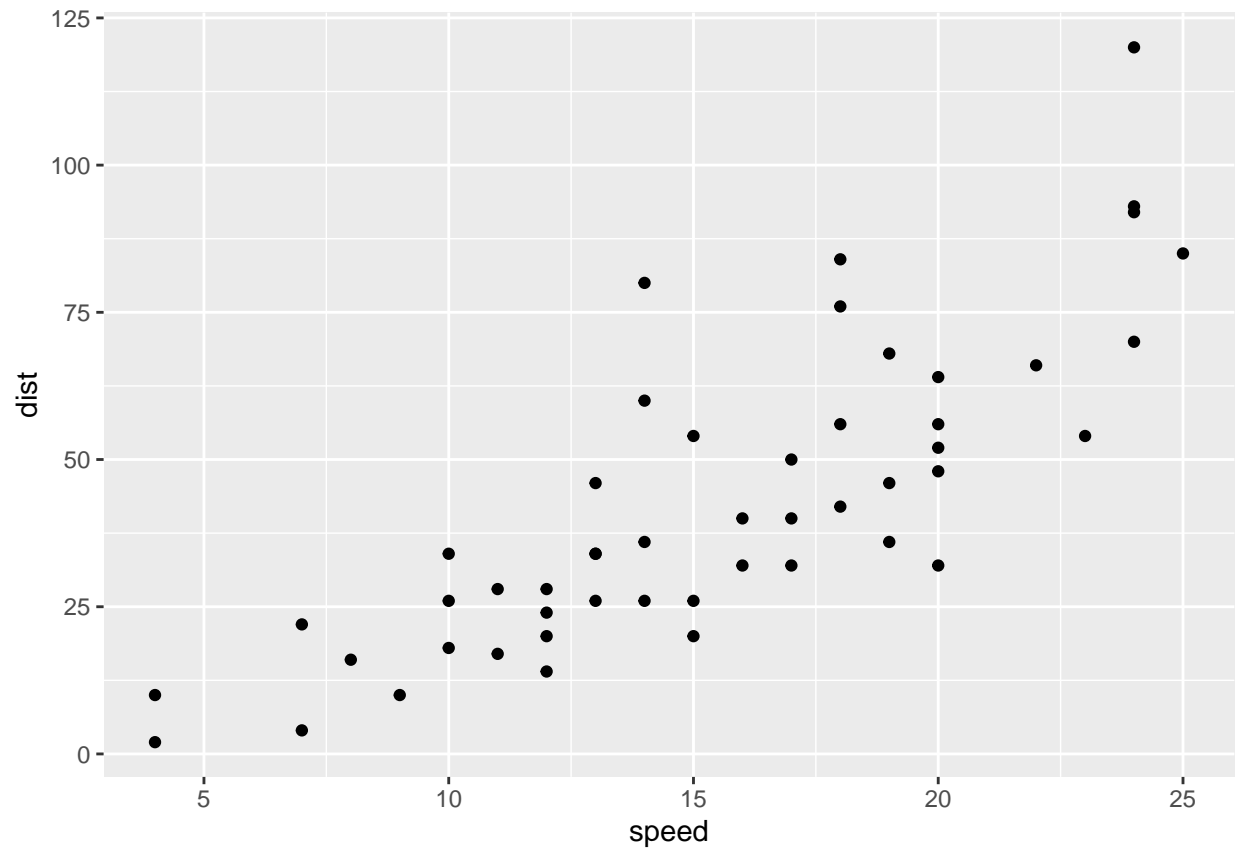
library(ggplot2)

View(cars)

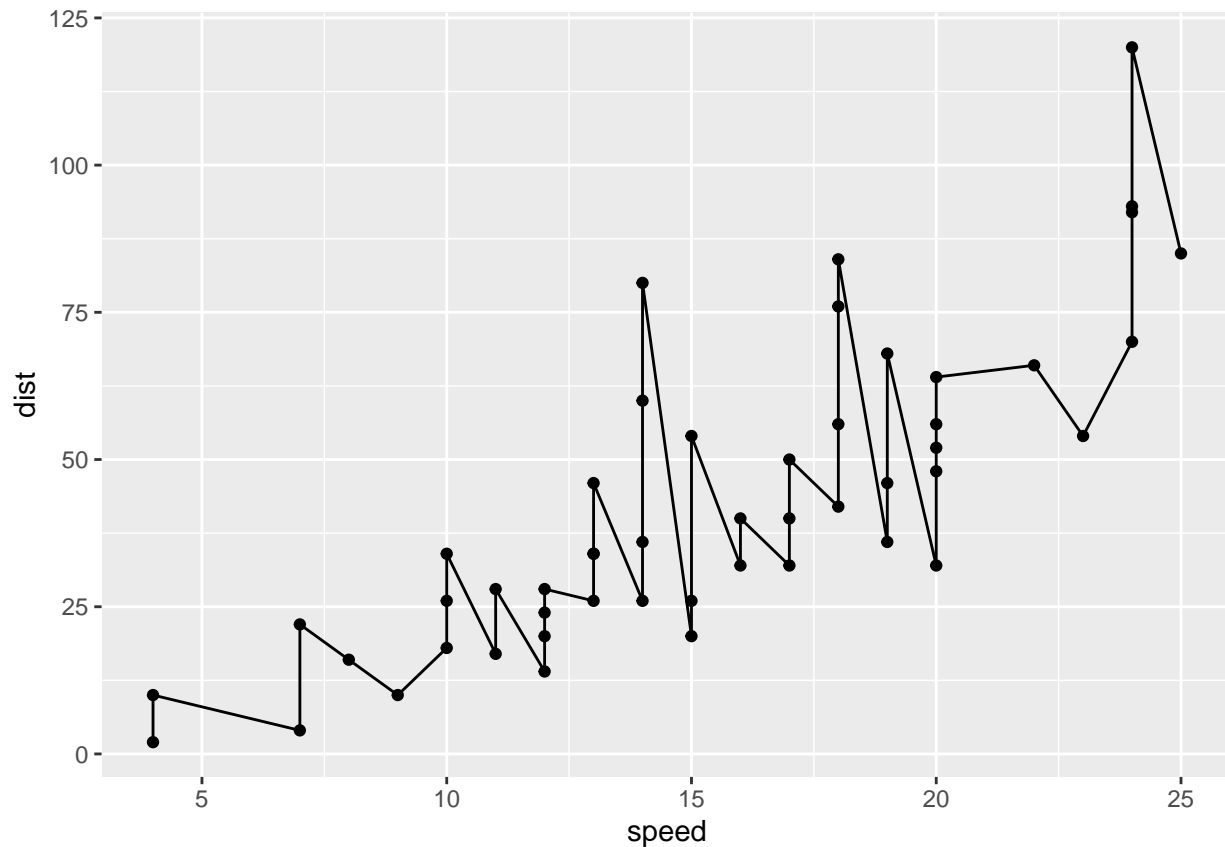
#Now we will use ggplot; we will need data + aes + geom

p <- ggplot(data = cars) +
  aes(x = speed, y = dist) +
  geom_point()

#regular scatter plot
p
```



```
#line plot  
p + geom_line()
```



#scatter plot with line

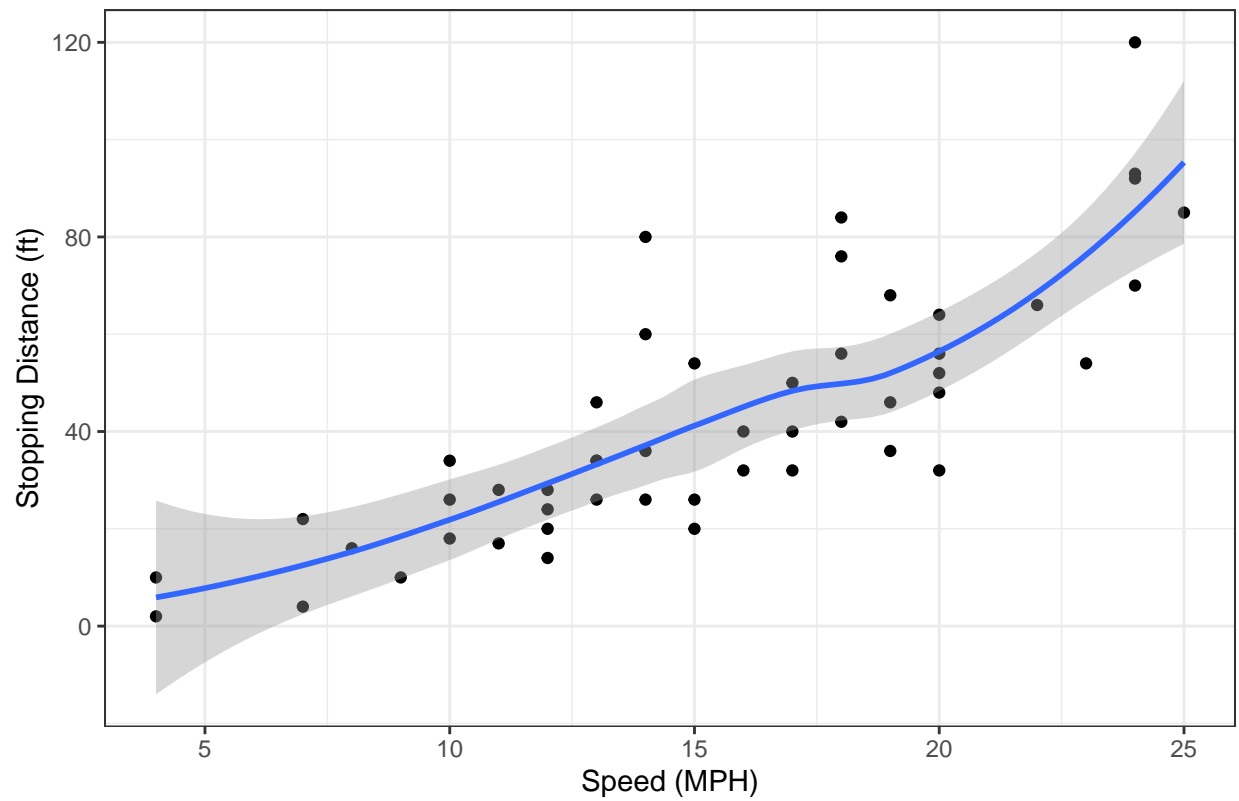
#Q6. In your own RStudio can you add a trend line layer to help show the relationship between the plot variables with the geom_smooth() function?

#Q7. Can you finish this plot by adding various label annotations with the labs() function and changing the plot look to a more conservative "black & white" theme by adding the theme_bw() function

```
p + geom_smooth() +  
  labs(x = "Speed (MPH)", y = "Stopping Distance (ft)",  
        title = "My First Nice ggplot") +  
  theme_bw()
```

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

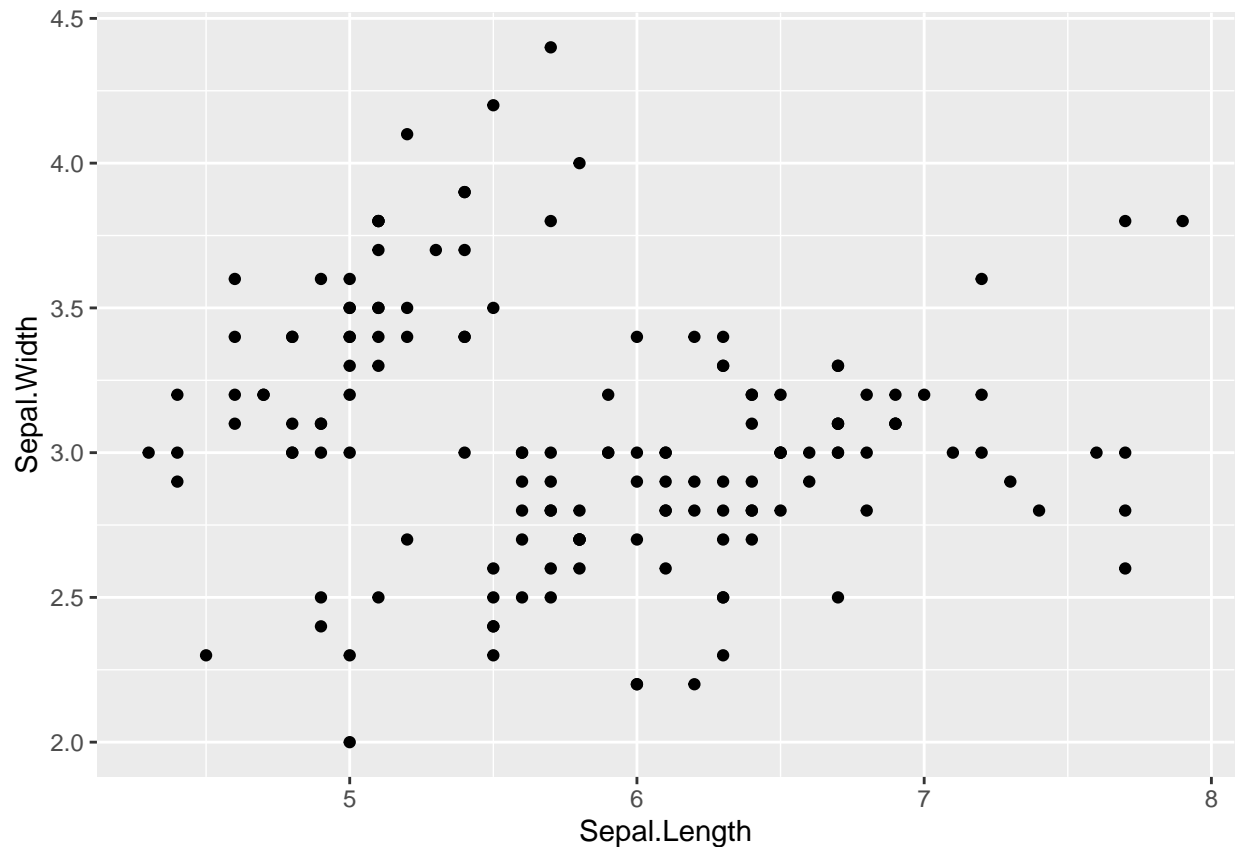
My First Nice ggplot



#Data-to-viz is a good place to get plot examples and their code
#<https://www.data-to-viz.com/>

#Basic Scatterplot

```
ggplot(iris, aes(x = Sepal.Length, y = Sepal.Width)) +  
  geom_point()
```



```
#Gene expression dataset
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
```

```
#Q8. Use the nrow() function to find out how many genes are in this dataset. What is your answer?
#5196
nrow(genes)
```

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

```
#Q9. Use the colnames() function and the ncol() function on the genes data frame to find out what the
#column names are (we will need these later) and how many columns there are. How many columns did you
#find?
#4 columns; "Gene", "Condition1", "Condition2", "State"
colnames(genes)
```

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

#Q10. Use the table() function on the State column of this data.frame to find out how many 'up' regulated genes there are. What is your answer?

```
#127 upregulated genes  
table(genes$State)
```

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

#Q11. Using your values above and 2 significant figures. What fraction of total genes is up-regulated in this dataset?

```
#2.44  
round( (table(genes$State) / nrow(genes)) * 100, 2)
```

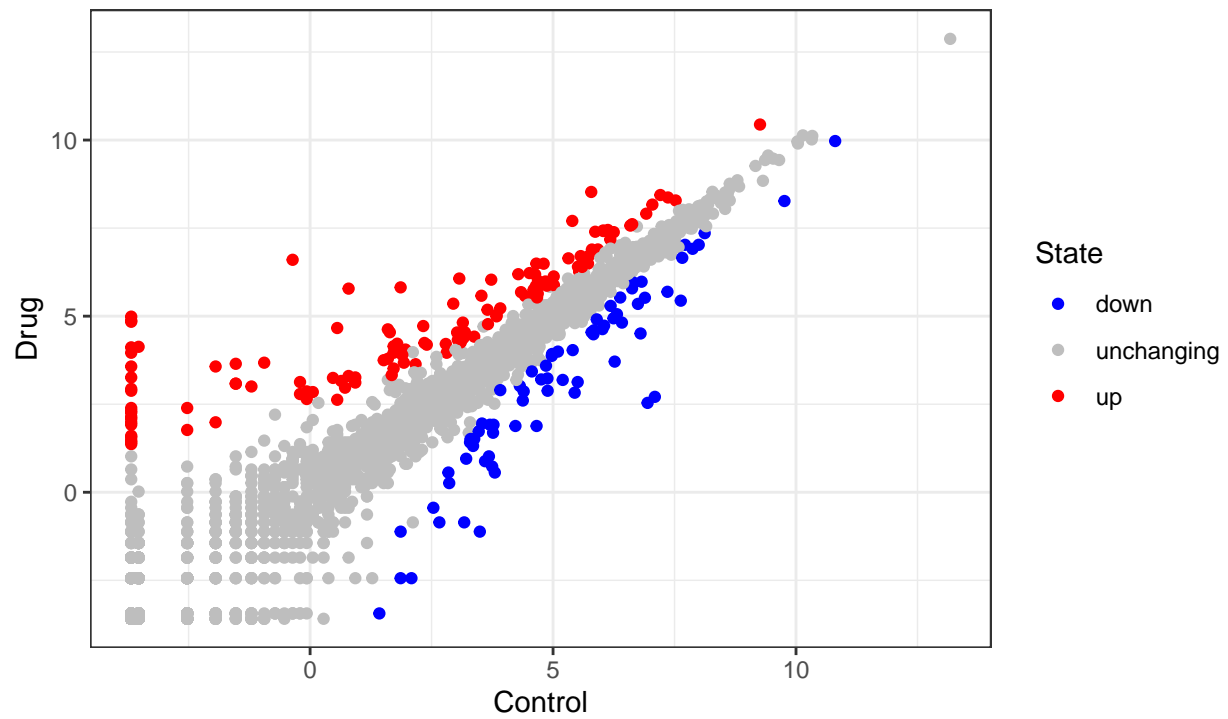
```
##  
##      down  unchanging      up  
##      1.39      96.17      2.44
```

#Making a figure the way I want it.

```
ggplot(data = genes) +  
  aes(x = Condition1, y = Condition2, color = State) +  
  geom_point() +  
  scale_colour_manual( values=c("blue","gray","red") ) +  
  theme_bw() +  
  labs(x = "Control", y = "Drug",  
       title = "Gene Expression Changes Upon Treatment",  
       subtitle = "Building in ggplot layer by layer",  
       caption = "Made by Ramola Baviskar")
```

Gene Expression Changes Upon Treatment

Building in ggplot layer by layer



Made by Ramola Baviskar