

# Week 4 Data visualization lab

Kalle Liimatta (PID A59002114)

October 12, 2022

```
# Week 4 Data visualization lab
```

```
View(cars)
```

```
#Simple base R plot
```

```
plot(cars)
```

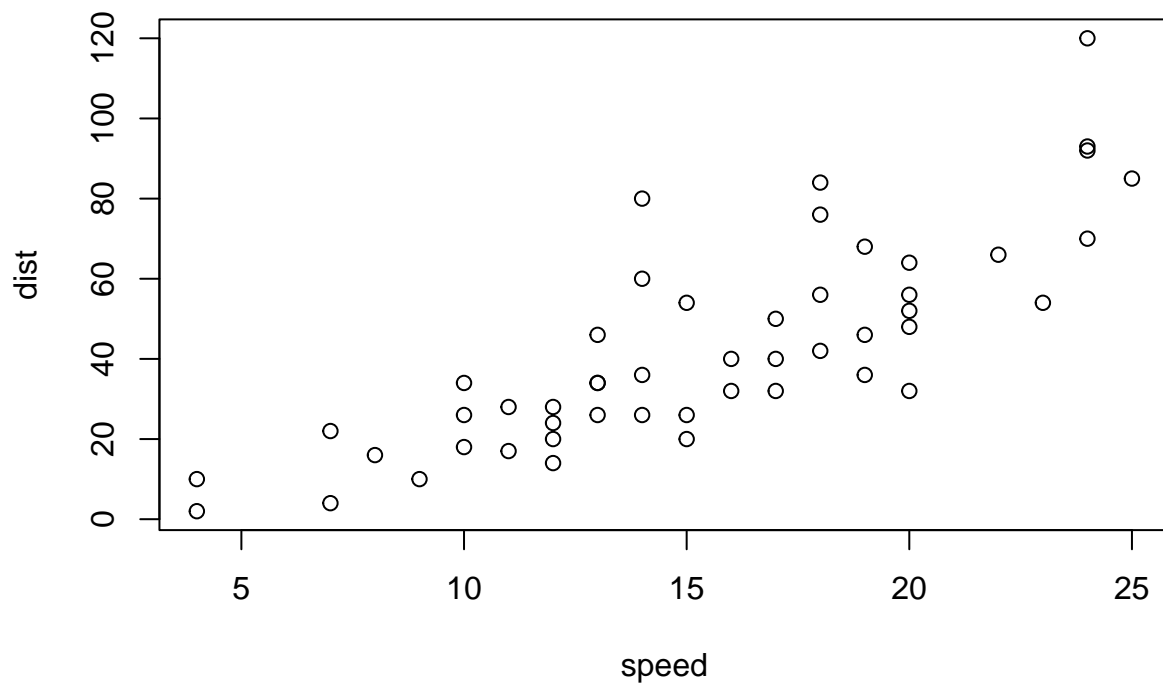
```
#One-time only install
```

```
#install.packages("ggplot2")
```

```
#Everytime you want to use a package in a
```

```
#new session, you have to load it with:
```

```
library(ggplot2)
```

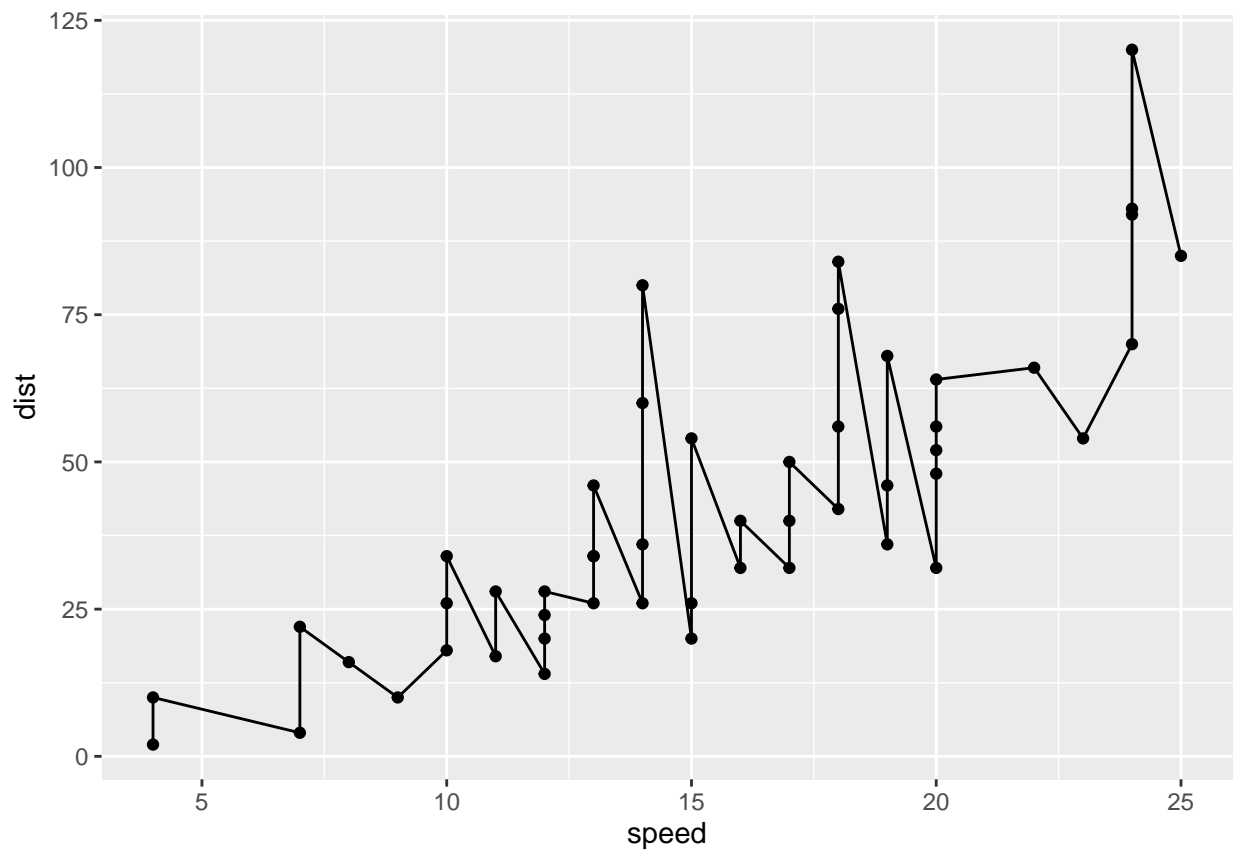


```

#Our first ggplot:
#We need data + aes + geom
p <- ggplot(cars) +
  aes(x=speed, y=dist) +
  geom_point()

#A silly line and point plot
p + geom_line()

```



```

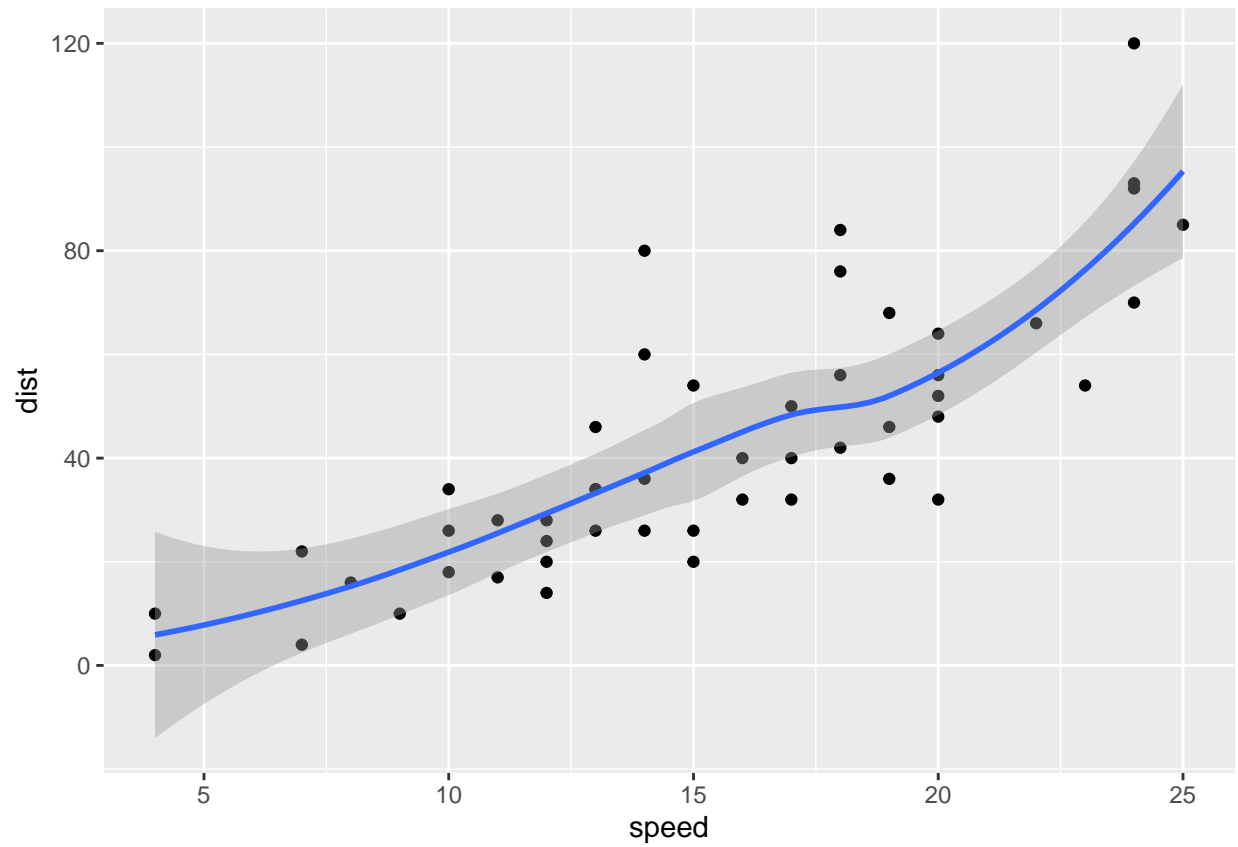
#Add a fitted line
p + geom_smooth()

```

```

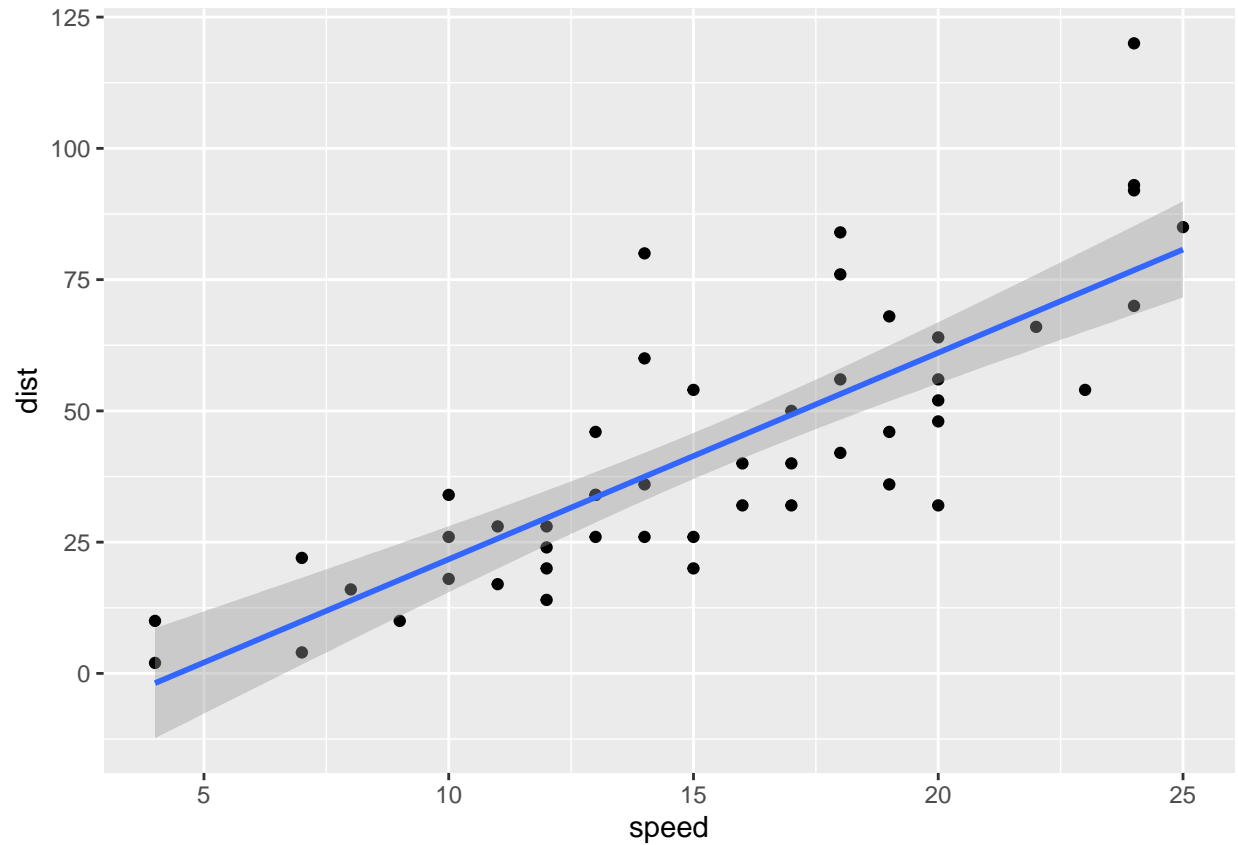
## 'geom_smooth()' using method = 'loess' and formula 'y ~ x'

```



```
#Fit a linear model  
p + geom_smooth(method="lm")
```

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

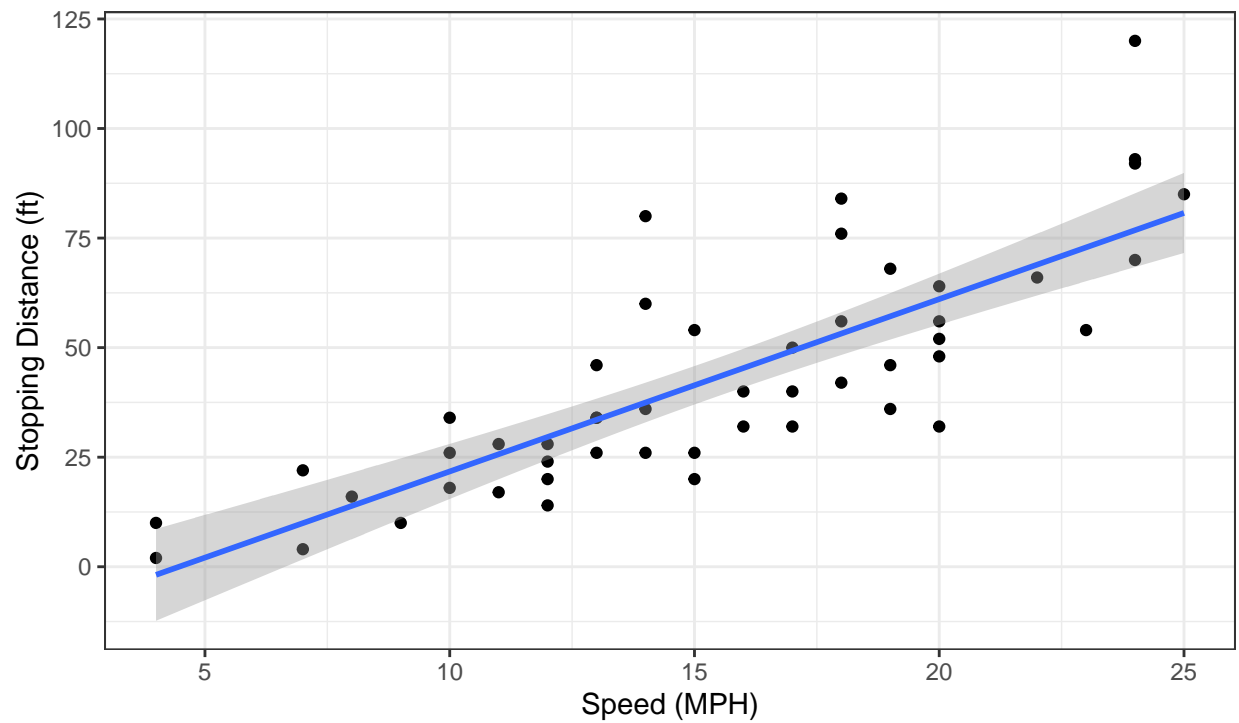


```
p + geom_smooth(method="lm") +  
  labs(title="Speed and Stopping Distances of Cars",  
        x="Speed (MPH)",  
        y="Stopping Distance (ft)",  
        subtitle = "Vroomin' and Zoomin'",  
        caption="Dataset: 'cars'") +  
  theme_bw()
```

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

## Speed and Stopping Distances of Cars

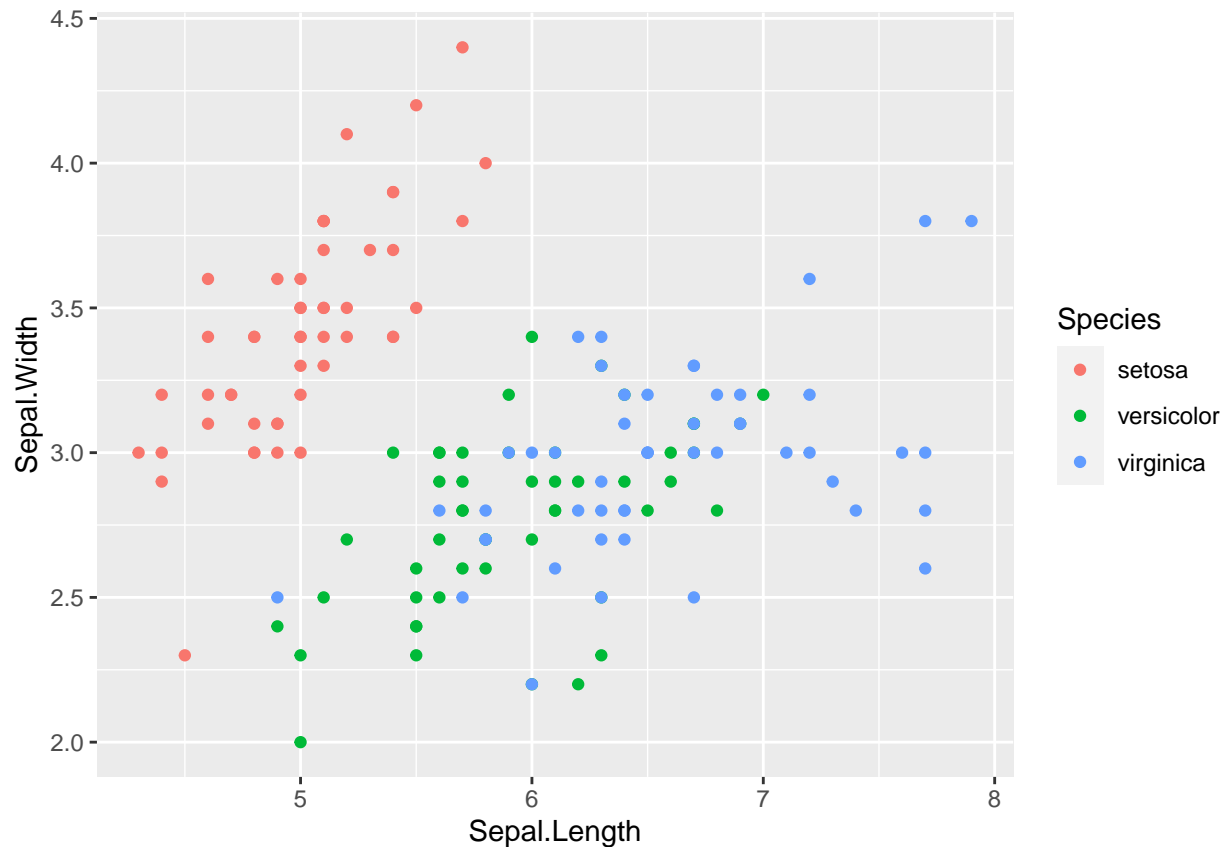
Vroomin' and Zoomin'



Dataset: 'cars'

*#Example plot from data2viz website*

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



```
#Make a more bioinformatic style plot
#RNAseq dataset for drug vs no drug treatment
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
```

```
#look at first 6 genes
View(genes)
```

```
#Q. how many genes are in this dataset
nrow(genes)
```

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

```
colnames(genes)
```

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

```
ncol(genes)
```

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

```
#Q. how many upregulated genes are there  
table(genes$State)
```

```
##  
##      down  unchanged      up  
##      72      4997      127
```

```
#Q. What fraction of genes are up/down/etc.  
round((table(genes$State)/nrow(genes))*100, 2)
```

```
##  
##      down  unchanged      up  
##      1.39      96.17      2.44
```

```
#Generate a publication figure
```

```
g <- ggplot(genes) +  
  aes(x=Condition1,  
      y=Condition2,  
      col=State) +  
  geom_point()  
  
#Add more custom layers  
g + theme_bw() +  
  scale_color_manual(values=c("blue", "gray", "red")) +  
  labs(x="Control (no drug)",  
       y="Drug treated",  
       title="Gene expression changes example plot")
```

Gene expression changes example plot

