

# Problem Set 10

Geoffrey Hughes

12/6/2019

a. Load Data

```
library("tidyverse")
```

```
## — Attaching packages —————  
————— tidyverse 1.2.1 —
```

```
## ✓ ggplot2 3.2.1      ✓ purrr  0.3.3  
## ✓ tibble  2.1.3      ✓ dplyr  0.8.3  
## ✓ tidyr   1.0.0      ✓ stringr 1.4.0  
## ✓ readr   1.3.1      ✓ forcats 0.4.0
```

```
## — Conflicts —————  
————— tidyverse_conflicts() —  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag()     masks stats::lag()
```

```
library("gapminder")  
library("gganimate")
```

```
setwd("/Users/geoffreyhughes/Documents/MGSC_310/MGSC310/Datasets")  
getwd()
```

```
## [1] "/Users/geoffreyhughes/Documents/MGSC_310/MGSC310/Datasets"
```

```
Scotch_DF <- read.csv("Scotch_DF.csv")
```

```
View(Scotch_DF)
```

```
dictionary <- data.frame(varlabel = c("yellow color", "amber color",
"peaty when smelled", "sweet when smelled", "fresh when smelled",
"fruit when smelled", "smooth texture", "smoky taste", "smooth taste",
"sweet taste", "fruit taste", "age of whiskey", "dist", "score (rating) of whiskey",
"region ", "district", "fraction alcohol", "islay ", "midland",
"spey", "east", "west", "north", "lowland", "campbell", "islands"),
varnan = c("color.yellow", "color.amber", "nose.peat", "nose.sweet",
"nose.fresh", "nose.fruit", "body.smooth", "pal.smoke",
"pal.smooth", "pal.sweet", "pal.fruit", "age", "dist",
"score", "region", "district", "percent", "islay", "midland",
"spey", "east", "west", "north", "lowland", "campbell",
"islands"))

print(dictionary)
```

```
##           varlabel      varnan
## 1      yellow color color.yellow
## 2      amber color  color.amber
## 3    peaty when smelled   nose.peat
## 4    sweet when smelled   nose.sweet
## 5    fresh when smelled   nose.fresh
## 6    fruit when smelled   nose.fruit
## 7      smooth texture body.smooth
## 8      smoky taste    pal.smoke
## 9      smooth taste    pal.smooth
## 10     sweet taste     pal.sweet
## 11     fruit taste     pal.fruit
## 12     age of whiskey      age
## 13           dist        dist
## 14 score (rating) of whiskey    score
## 15           region      region
## 16     district        district
## 17   fraction alcohol    percent
## 18           islay        islay
## 19     midland        midland
## 20           spey        spey
## 21           east        east
## 22           west        west
## 23           north        north
## 24     lowland        lowland
## 25     campbell    campbell
## 26     islands        islands
```

#### b. Familiarize & Factor

- I would factor those variables that aren't in binary format; so the other ints, doubles, and strings.
- I do this so that those variables can be used as a number in training models / other math stuff
- Factor: age, dist, score, region, district, percent

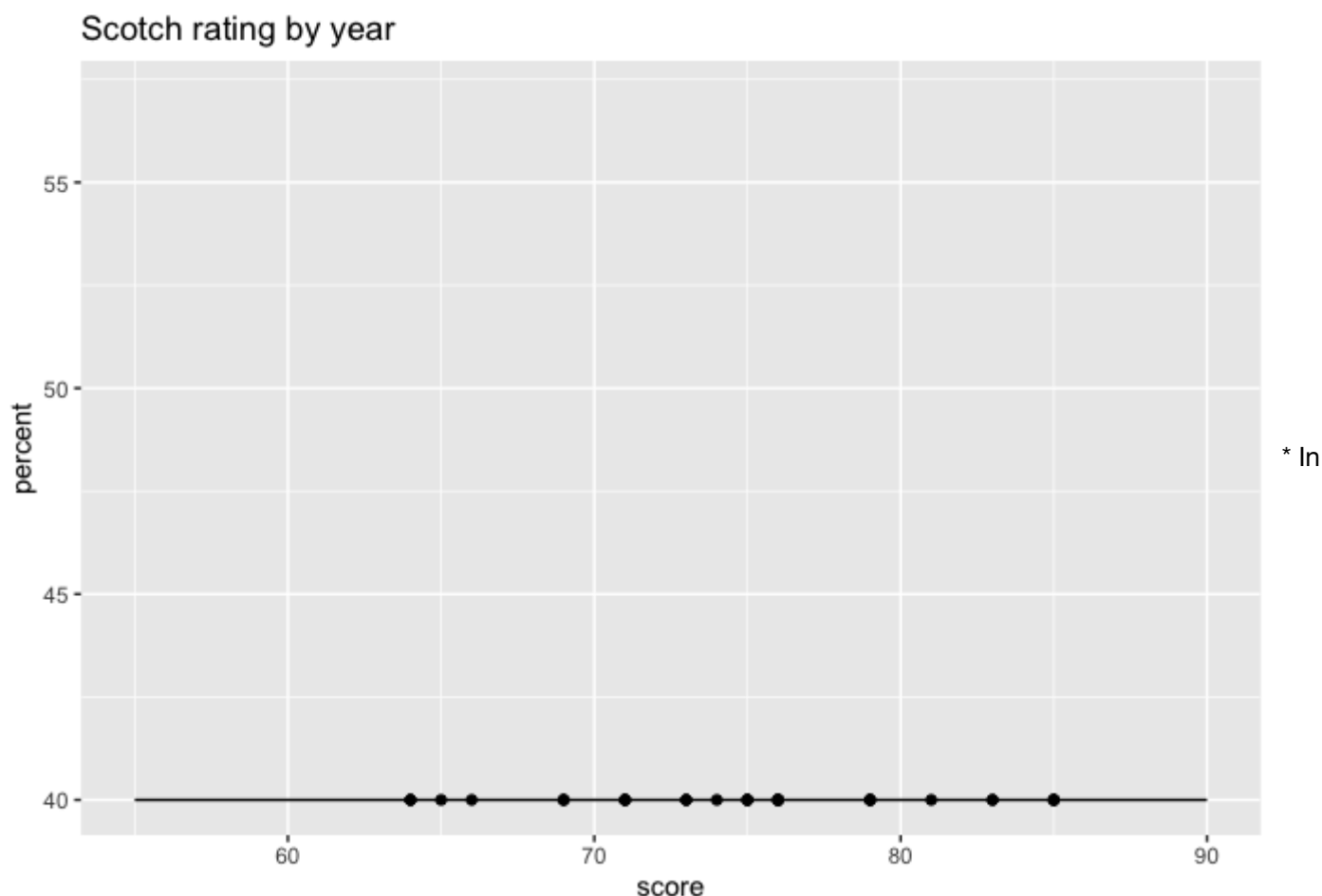
#### c. 2 plots showing how score varies with other variables

```
q <- ggplot(Scotch_DF, aes(x = score, y = age)) +
  geom_line() +
  geom_point() +
  transition_reveal(age) +
  ease_aes('linear') +
  ggtitle("Scotch rating by year") +
  theme(legend.position = 'none')

animate(q, 100, 20)

p <- ggplot(Scotch_DF, aes(x = score, y = percent)) +
  geom_line() +
  geom_point() +
  transition_reveal(percent) +
  ease_aes('linear') +
  ggtitle("Scotch rating by year") +
  theme(legend.position = 'none')

animate(p, 100, 20)
```



the first graph we see some peak scores around 10-15 years, and then a dropoff from 80+ score at 15 to less than 60 at 20 years. It seems scotch peaks in score around 10-15 years, and drastically lowers in score after 15 years.

- From the second graph we can tell that not many Scotches are more than 45-50% alcohol (ABV). Also the only few scotches that make it past 55% rebound on score, and converge to about score = 80.