

Grand Slam

1. Load packages

```
library (tidyverse)
library (janitor)
library (ggthemes)
library (ggrepel)
library (stringr)
```

2. Source data

```
tennis_raw <- read_csv("https://raw.githubusercontent.com/rfordatascience/tidyuesday/master/data/2019/2019-Grand-Slam-tennis.csv")
```

3. Convert outcomes to numeric

```
tennis <- tennis_raw %>%
  clean_names() %>%
  mutate(
    outcome_num = case_when(
      outcome == "Won" ~ 8,
      outcome == "Finalist" ~ 7,
      outcome == "Semi-finalist" ~ 6,
      outcome == "Quarterfinalist" ~ 5,
      outcome == "4th Round" ~ 4,
      outcome == "3rd Round" ~ 3,
      outcome == "2nd Round" ~ 2,
      outcome == "1st Round" ~ 1
    ),
    player = str_remove(player, "// "),
    player = str_replace(player, "Seleš", "Seles")
  ) %>%
  drop_na()

tennis %>%
  head(10)
```

```
## # A tibble: 10 x 6
##   player          year tournament outcome gender outcome_num
##   <chr>          <dbl> <chr>      <chr>    <chr>      <dbl>
## 1 Margaret Court  1968 Australian 0~ Finalist  Female          7
## 2 Billie Jean Moffitt~ 1968 Australian 0~ Won       Female          8
## 3 Lesley Turner Bowrey 1968 Australian 0~ Semi-finali~ Female          6
## 4 Rosemary Casals     1968 Australian 0~ Quarterfina~ Female          5
## 5 Judy Tegart Dalton   1968 Australian 0~ Semi-finali~ Female          6
## 6 Kerry Melville Reid  1968 Australian 0~ 3rd Round  Female          3
## 7 Evonne Goolagong Ca~ 1968 Australian 0~ 3rd Round  Female          3
```

```
## 8 Lorraine Coghlan Ro~ 1968 Australian O~ 1st Round Female 1
## 9 Helen Gourlay Cawley 1968 Australian O~ 1st Round Female 1
## 10 Billie Jean Moffitt~ 1968 French Open Semi-finali~ Female 6
```

4. Calculate average outcomes

```
tennis_avg <- tennis %>%
  group_by (player) %>%
  summarize (
    avg = round(mean (outcome_num),2),
    st_dev = round(sd (outcome_num),2),
    n = n ()
  ) %>%
  arrange (-avg) %>%
  filter (n>9) %>%
  drop_na ()

tennis_avg %>%
  head (10)
```

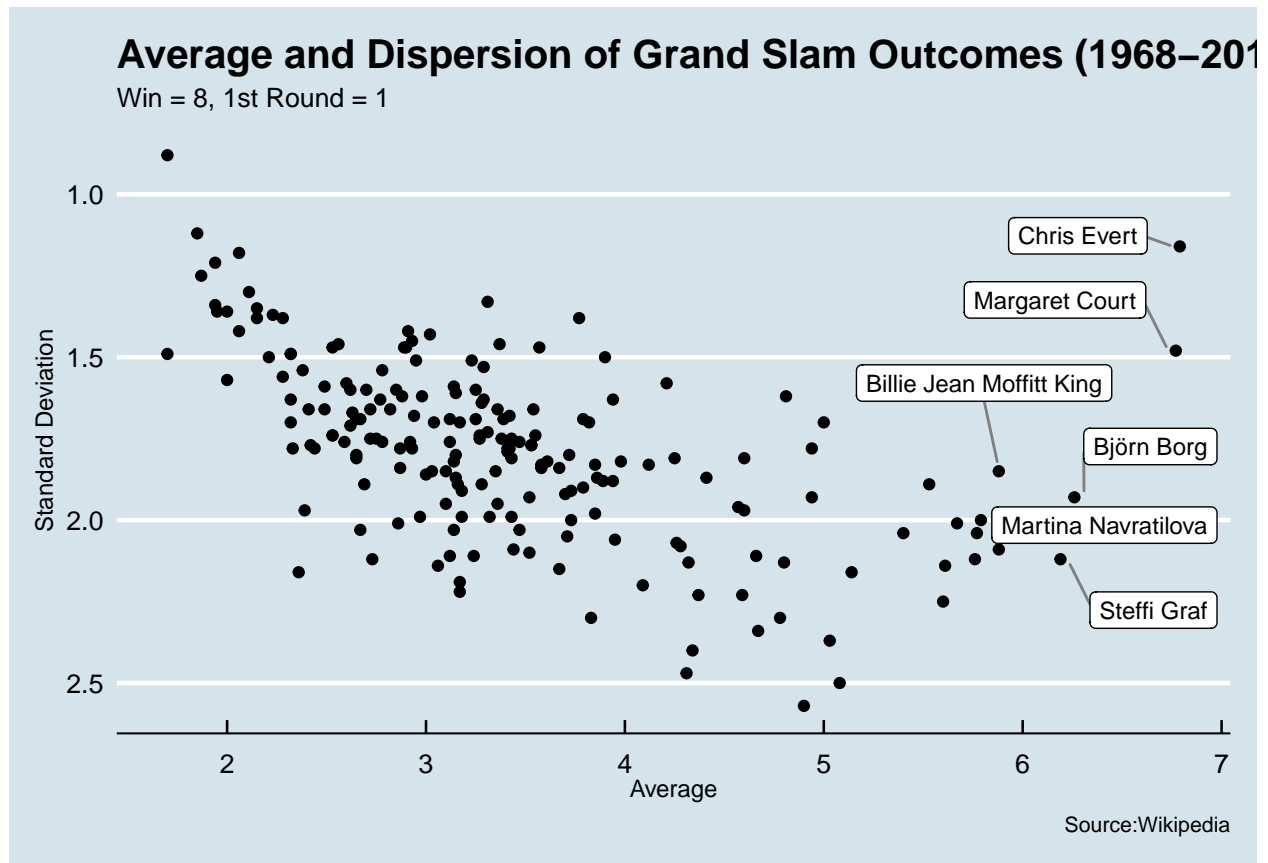
```
## # A tibble: 10 x 4
##   player          avg st_dev    n
##   <chr>         <dbl> <dbl> <int>
## 1 Chris Evert      6.79  1.16  56
## 2 Margaret Court   6.77  1.48  22
## 3 Björn Borg       6.26  1.93  27
## 4 Steffi Graf      6.19  2.12  54
## 5 Billie Jean Moffitt King 5.88  1.85  33
## 6 Martina Navratilova 5.88  2.09  67
## 7 Serena Williams  5.79  2    70
## 8 Novak Djokovic   5.77  2.04  56
## 9 Evonne Goolagong Cawley 5.76  2.12  34
## 10 Ivan Lendl      5.67  2.01  42
```

4. Create visualization

```
tennis_plot <- ggplot (tennis_avg, aes (avg, st_dev, label = player)) +
  geom_point () +
  theme_economist() +
  scale_y_continuous(trans = "reverse") +
  geom_label_repel(
    aes(label=ifelse(avg>5.8,as.character(player),'')),
    box.padding = 0.35,
    point.padding = 0.5,
    size = 3,
    segment.color = 'grey50') +
  labs(
    title = "Average and Dispersion of Grand Slam Outcomes (1968-2018)",
    subtitle = "Win = 8, 1st Round = 1",
    caption = "Source:Wikipedia",
```

```
x = "Average",
y = "Standard Deviation"
) +
theme(
  axis.title.x=element_text(size=9),
  axis.title.y=element_text(size=9)
)
```

tennis_plot



```
ggsave ("tennis.png", tennis_plot)
```

Saving 6.5 x 4.5 in image