Data Viz Recreation Project

Amelie Devine

2024-03-05

Download packages & data

```r
library(tidyverse)
```

```
## ── Attaching core tidyverse packages ──────────────────────── tidyverse
2.0.0 ──
## ✓ dplyr     1.1.2     ✓ readr     2.1.4
## ✓ forcats   1.0.0     ✓ stringr   1.5.0
## ✓ ggplot2   3.5.1     ✓ tibble    3.2.1
## ✓ lubridate 1.9.3     ✓ tidyr     1.3.0
## ✓ purrr     1.0.2
## ── Conflicts ────────────────────────────────────────────
tidyverse_conflicts() ──
## ✗ dplyr::filter() masks stats::filter()
## ✗ dplyr::lag()    masks stats::lag()
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all
conflicts to become errors
```

```r
data <- read_csv("NBA_Salary_History.csv", show_col_types = FALSE)
```

summary of data

```r
data
```

```
## # A tibble: 887 × 4
##    Season  Team                `Salary Cap` `Total Salary`
##    <chr>   <chr>                      <dbl>          <dbl>
##  1 1990-91 Atlanta Hawks           11871000       11761000
##  2 1990-91 Boston Celtics          11871000       11256000
##  3 1990-91 Charlotte Hornets       11871000       10417000
##  4 1990-91 Chicago Bulls           11871000       10040000
##  5 1990-91 Cleveland Cavaliers     11871000       14403000
##  6 1990-91 Dallas Mavericks        11871000       11693000
##  7 1990-91 Denver Nuggets          11871000       10335000
##  8 1990-91 Detroit Pistons         11871000       12910000
##  9 1990-91 Golden State Warriors   11871000       11150000
## 10 1990-91 Houston Rockets         11871000       10500000
## # ℹ 877 more rows
```

```r
summary(data)
```

```
##     Season              Team              Salary Cap        Total Salary
##  Length:887         Length:887         Min.   : 4233000   Min.   :
```

```
2914500
##  Class :character   Class :character   1st Qu.:23000000   1st Qu.:
23924289
##  Mode  :character   Mode  :character   Median :43840000   Median :
54725500
##                                        Mean   :42092849   Mean   :
50672603
##                                        3rd Qu.:58044000   3rd Qu.:
69943381
##                                        Max.   :99093000   Max.
:137494845
```

Modify salary data to be in millions of dollars

```r
data$`Salary_Cap` <- (data$`Salary Cap` / 1000000)
data$`Total_Salary` <- (data$`Total Salary` / 1000000)
```

Modify data to find max, min, average for Total_Salary for each season

```r
data2 <- data %>%
    group_by(Season) %>%
    summarize(min_salary = min(Total_Salary),
              max_salary = max(Total_Salary),
              avg_salary = mean(Total_Salary),
              salary_cap = unique(Salary_Cap)) %>%
    filter(Season != "2017-18")
```

Smooth the curves

```r
spline_int_min <- as.data.frame(spline(row.names(data2), data2$min_salary))
spline_int_max <- as.data.frame(spline(row.names(data2), data2$max_salary))
spline_int_avg <- as.data.frame(spline(row.names(data2), data2$avg_salary))

ggplot(data2, aes(x = Season)) +
    geom_bar(stat = "identity", width = 0.6, fill = "#f6d28d",
             aes(y = salary_cap, color = "Salary Cap"), size = 1) +
    geom_line(data = spline_int_min, aes(x = x, y = y, color = "Minimum
Salary"),
              linewidth = 0.85) +
    geom_line(data = spline_int_max, aes(x = x, y = y, color = "Maximum
Salary"),
              linewidth = 0.85) +
    geom_line(data = spline_int_avg, aes(x = x, y = y, color = "Average
Salary"),
              linewidth = 0.85) +
    geom_point(aes(y = min_salary, color = "Minimum Salary"),
               size = 2, shape = 1, stroke = 1.25) +
    geom_point(aes(y = max_salary, color = "Maximum Salary"),
               size = 2, shape = 1, stroke = 1.25) +
    geom_point(aes(y = avg_salary, color = "Average Salary"),
               size = 2, shape = 1, stroke = 1.25) +
    labs(title = "Historical Team Spending against the Cap in the Salary Cap
```

```r
Era",
       subtitle = "1985/86 - 2016/17: Not adjusted for inflation.",
       y = "USD (in millions)",
       x = NULL) +
  scale_y_continuous(breaks = seq(0, 140, by = 20),
                     labels = c("$0", "$20", "$40", "$60", "$80", "$100",
                                "$120", "$140")) +
  scale_x_discrete(breaks = data2$Season[c(TRUE, FALSE, TRUE, FALSE, TRUE,
FALSE,
                                           TRUE, FALSE, TRUE, FALSE, TRUE,
FALSE,
                                           TRUE, FALSE, TRUE, FALSE, TRUE,
FALSE,
                                           TRUE, FALSE, TRUE, FALSE, TRUE,
FALSE,
                                           TRUE, FALSE, TRUE, FALSE, FALSE,
TRUE)],
                   labels = c("1985-86", "1988-89", "1991-92", "1993-94",
                              "1995-96", "1997-98", "1999-00", "2001-02",
                              "2003-04", "2005-06", "2007-08", "2009-10",
                              "2011-12", "2013-14", "2016-17")) +
  coord_cartesian(ylim = c(0, 140)) +
  scale_color_manual(name = "",
                     values = c("Maximum Salary" = "#0f6db4",
                                "Average Salary" = "black",
                                "Minimum Salary" = "#ce173e",
                                "Salary Cap" = "#f1b541"),
                     labels = c("Highest Payroll (MAX)",
                                "Average Payroll (AVG)",
                                "Lowest Payroll (MIN)",
                                "Salary Cap (CAP)")) +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1.2, vjust = 1.6),
        axis.ticks = element_blank(),
        plot.title = element_text(hjust = 0.5),
        plot.subtitle = element_text(hjust = 0.5),
        panel.grid.major.x = element_blank(),
        panel.grid.minor.x = element_blank(),
        panel.grid.minor.y = element_blank(),
        panel.border = element_blank(),
        legend.position = "bottom")

## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

Historical Team Spending against the Cap in the Salary Ca

1985/86 - 2016/17: Not adjusted for inflation.

Highest Payroll (MAX)   —□— Average Payroll (AVG)   —□— Lowest Payroll (MIN)   ▢   Sa