SAL 413 HW4

Hunter Geise

2023-11-03

R Markdown

Overview and Instructions This problem set is intended to exercise your coding skills. Your styling must conform to the tidyverse style guide. Make sure your code, output, and responses are all present. You must work alone on this assignment. However, if you find yourself stuck, you may seek help from your peers. You should not have the exact same code as a classmate. You may use whatever notes, textbooks, etc. you find helpful for the assignment. Your code should consist of functions found in the tidyverse packages we have gone over in class to this point. Only use Base R functions if there is no other way to accomplish a task within the tidyverse. Please compile your document into a PDF, and upload the PDF to Blackboard with the filename in the following format: [surname]_hw4.pdf. For example, I would turn in maddox_hw4.pdf. Please type out the questions on your turn-in. Please start each new question on a new page.

Questions This homework will work on data manipulation primarily using dplyr, but other tidyverse packages may also be used. Each question will build on the results from the previous question. Read each question carefully as steps to solving the question may be laid out in the question itself. Download the nfl_team_revenue_data_combined.csv dataset from Blackboard. Use the readr package when doing so to ensure the data is read in as a tibble. The data has revenue data for each NFL team from the 2001 – 2016 seasons, along with the division that the team is currently in for the 2023 season. For this assignment, assume that the division teams are in currently is the same division the team has been in since 2001.

```
library(tidyverse)
library(purrr)
library(dplyr)
library(readr)
library(stringr)
theme_set(theme_bw())
library(tibble)
library(ggplot2)
```

Question 1: For each season in the data (2001 – 2016) find the division that averaged the largest revenue. Create a table that includes division, season, and avg_rev that just has the top division listed for each season (only 16 rows in table). Order the table by season.

```
data <- read_csv("hw_4.csv")
```

```
## Rows: 510 Columns: 4
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (2): Team, Division (Currnet)
## dbl (2): Season, Revenue
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
data %>% rename(Division = `Division (Currnet)`) -> data

data %>%
  group_by(Season, Division) %>%
  summarize(avg_rev = mean(Revenue)) %>%
  top_n(1, avg_rev) %>%
  ungroup()-> avg_revenue
```

```
## `summarise()` has grouped output by 'Season'. You can override using the
## `.groups` argument.
```

```
avg_revenue %>% arrange(Season) -> ordered_avg_revenue
ordered_avg_revenue
```

```
## # A tibble: 16 x 3
##    Season Division avg_rev
##     <dbl> <chr>      <dbl>
##  1   2001 NFC East    162.
##  2   2002 NFC East    176.
##  3   2003 NFC East    200.
##  4   2004 NFC East    227.
##  5   2005 NFC East    234.
##  6   2006 NFC East    243.
##  7   2007 NFC East    262.
##  8   2008 NFC East    276.
##  9   2009 NFC East    318.
## 10   2010 NFC East    331.
## 11   2011 NFC East    374.
## 12   2012 NFC East    391
## 13   2013 NFC East    410.
## 14   2014 NFC East    457.
## 15   2015 NFC East    500.
## 16   2016 NFC East    557.
```

Question 2: One team appears in this dataset two times less than the other teams. Write code to find which team that is.

```
data %>%
  group_by(Team) %>%
  count(Team) %>%
  arrange(n) %>%
  ungroup()-> team_tally

team_tally %>% head(3)
```

```
## # A tibble: 3 x 2
##   Team                  n
##   <chr>             <int>
## 1 Houston Texans       14
## 2 Arizona Cardinals    16
## 3 Atlanta Falcons      16
```

Question 3: Create a new variable within the data that standardizes (z-score) the revenue for each season based on the revenue generated by other teams during that season.

```
data %>%
  group_by(Season) %>%
  mutate(z_score = scale(Revenue)) %>%
  ungroup()-> revenue_scale

revenue_scale
```

```
## # A tibble: 510 x 5
##    Team              Season Revenue Division z_score[,1]
##    <chr>              <dbl>   <dbl> <chr>          <dbl>
##  1 Arizona Cardinals   2001     110 NFC West      -1.44
##  2 Arizona Cardinals   2002     126 NFC West      -1.28
##  3 Arizona Cardinals   2003     131 NFC West      -1.42
##  4 Arizona Cardinals   2004     153 NFC West      -1.13
##  5 Arizona Cardinals   2005     158 NFC West      -1.17
##  6 Arizona Cardinals   2006     189 NFC West      -0.578
##  7 Arizona Cardinals   2007     203 NFC West      -0.664
##  8 Arizona Cardinals   2008     223 NFC West      -0.481
##  9 Arizona Cardinals   2009     236 NFC West      -0.346
## 10 Arizona Cardinals   2010     240 NFC West      -0.542
## # i 500 more rows
```

Question 4: Plot the revenue for each team by season. Within this plot, create a different plot for each division, and organize the plot such that all NFC divisions are on the left and AFC division are on the right. Also have the divisions ordered from top to bottom as NESW. Therefore the NFC North will be the top left graph and the AFC West will be the bottom right graph. Within each of these subgraphs plot a different colored line for each team that shows their revenue by season. Overlay each divisional plot with the average revenue for the division in a color not used for any of the teams. Then overlay the plots with the overall NFL average revenue with a black line that is larger than the other lines.

```
data$Division <- factor(data$Division,
                        levels = c("NFC North", "AFC North",
                                   "NFC East", "AFC East",
                                   "NFC South", "AFC South",
                                   "NFC West", "AFC West"))

division_avg <- data %>%
  group_by(Season, Division) %>%
  summarize(avg_rev = mean(Revenue))
```

```
## `summarise()` has grouped output by 'Season'. You can override using the
## `.groups` argument.
```

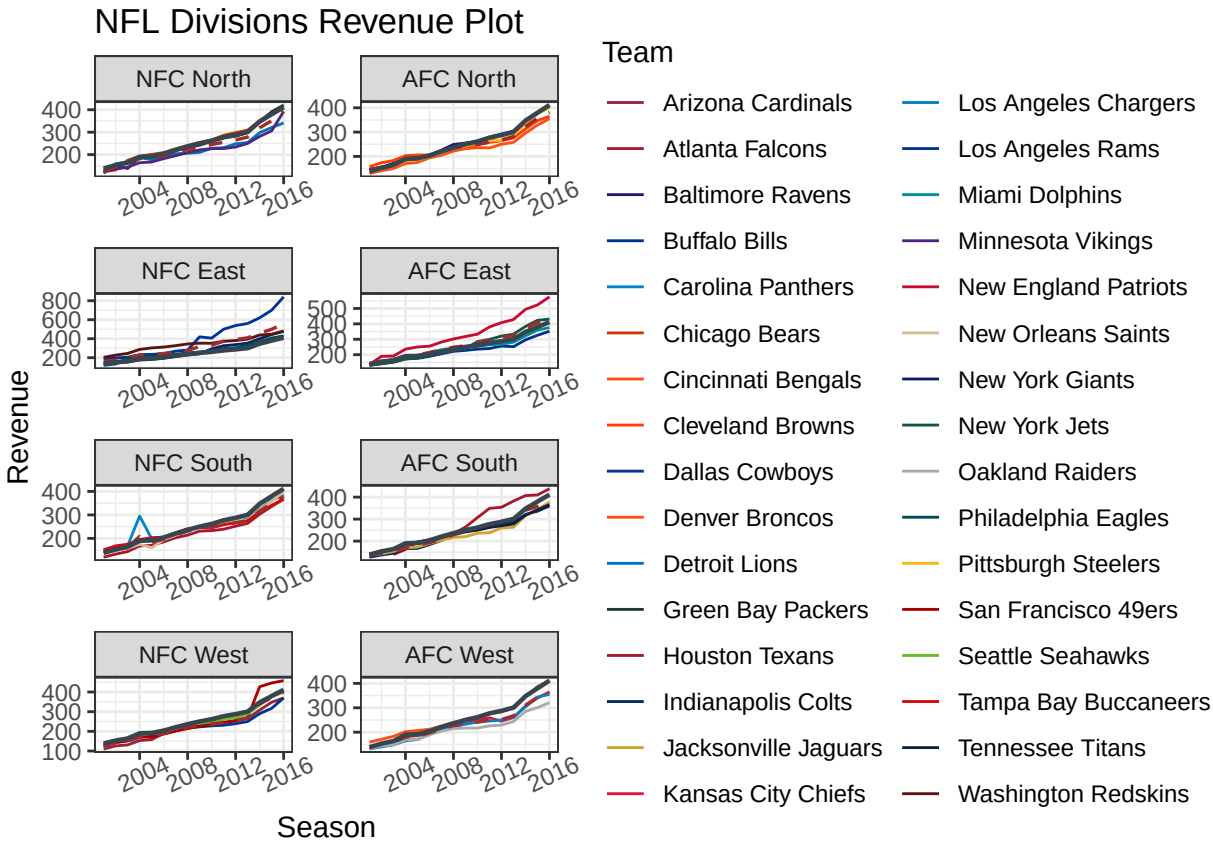
```
league_avg <- data %>%
  group_by(Season) %>%
  summarize(avg_rev = mean(Revenue))

p1 <- ggplot(data, aes(Season, Revenue)) +
  geom_line(aes(color = Team)) +
  facet_wrap(vars(Division), ncol = 2, scales = "free") +
  labs(title = "NFL Divisions Revenue Plot") +
  theme(axis.text.x = element_text(angle=25)) +
  scale_color_manual(values = c("#97233F", "#A71930", "#241773", "#00338D",
                                "#0085CA", "#C83803", "#FB4F14", "#FF3C00",
                                "#003594", "#FB4F14", "#0076B6", "#203731",
                                "#A71930", "#002C5F", "#D7A22A", "#E31837",
                                "#0080C6", "#003594", "#008E97", "#4F2683",
                                "#C60C30", "#D3BC8D", "#0B2265", "#125740",
                                "#A5ACAF", "#004C54", "#FFB612", "#AA0000",
                                "#69BE28", "#D50A0A", "#0C2340", "#5A1414"))

newp1 <- p1 +
  geom_line(data = division_avg, aes(y = avg_rev), color = "brown", linewidth = .6,
            linetype = "dashed", show.legend = FALSE)

finalp1 <- newp1 +
  geom_line(data = league_avg, aes(y = avg_rev), color = "#36454F", linewidth = .8,
            show.legend = FALSE)

finalp1
```


NFL Divisions Revenue Plot
NFC North
AFC North
NFC East
AFC East
NFC South
AFC South
NFC West
AFC West
Revenue
Season
2004
2008
2012
2016
Team
Arizona Cardinals
Atlanta Falcons
Baltimore Ravens
Buffalo Bills
Carolina Panthers
Chicago Bears
Cincinnati Bengals
Cleveland Browns
Dallas Cowboys
Denver Broncos
Detroit Lions
Green Bay Packers
Houston Texans
Indianapolis Colts
Jacksonville Jaguars
Kansas City Chiefs
Los Angeles Chargers
Los Angeles Rams
Miami Dolphins
Minnesota Vikings
New England Patriots
New Orleans Saints
New York Giants
New York Jets
Oakland Raiders
Philadelphia Eagles
Pittsburgh Steelers
San Francisco 49ers
Seattle Seahawks
Tampa Bay Buccaneers
Tennessee Titans
Washington Redskins