Friends Data Analytics Analysis

David Cohn

6/27/2021

I: Background

The sitcom Friends, originally airing on NBC from 1994-2004, was one of the most successful TV shows of the 1990's (along with Seinfeld and ER). Even after its original run, Friends continues to reach old and new audiences on syndication and streaming, as evidenced by recent interest in the Friends Reunion special on HBO Max, starring the original six members of its ensemble cast: Jennifer Anniston, Courteney Cox, Lisa Kudrow, Matt LeBlanc, Matthew Perry, and David Schwimmer.

Given the significant recent interest in the Friends Reunion show, I wanted to perform an analysis to better understand quantitatively how Friends has achieved such sustained success over 25+ years and 10 seasons.

II: Dataset and Processing

Wikipedia, as part of a summary entry on Friends, compiled publically available season, episode number, title, writer, director, original air date and Nielsen viewership data for all 236 episode of Friends across 10 seasons (Link: https://en.wikipedia.org/w/index.php?title=List_of_Friends_episodes&oldid=1029964325); the first nine seasons of Friends generally averaged either 24 or 25 episodes, while the tenth season was an abbreviated 18 episode season. In addition, IMDB has publically available viewer ratings data for each Friend episode (Link: https://www.imdb.com/title/tt0108778/?ref_=tt_ov_inf), based on a 1-10 scale. I compiled this Wikipedia and IMDB data in a .csv file, and proceeded to split each episode's writer credits into a series of columns, with one column per writer. In making an assumption for the following analysis, I treated story and teleplay writer credits (if applicable) equally.

```r
# R Packages Used In Analysis
library(ggplot2)
library(dplyr)
library(tidyr)
library(lubridate)
```

```r
friends.database = filter(friends.database, !is.na(Season))

# Remove Quotation Marks from Episode Titles
friends.database$Title = gsub('"', '', friends.database$Title, fixed = TRUE)

friends.directors = data.frame(unique(friends.database$Director))
names(friends.directors) = c("Director")

friends.database$Air.Date = as.character(friends.database$Air.Date)
friends.database$Air.Date = as.Date(friends.database$Air.Date, format = '%m/%d/%y')

friends.database$Season = as.character(friends.database$Season)
friends.database$Season = factor(friends.database$Season, levels = seq(1, 10))
```

```r
friends.database$Episode = as.character(friends.database$Episode)
friends.database$Episode = factor(friends.database$Episode, levels = seq(1, 25))

friends.database = friends.database %>%
  # Calculate Month of Air Date (with Month Abbreviation Label)
  mutate(Month = month(Air.Date, label = TRUE)) %>%
  mutate(Year = year(Air.Date)) %>%
  mutate(Episode.Number = row.names(friends.database))

friends.database$Month = as.factor(friends.database$Month)
```

III: Analysis

From reading Wikipedia articles on Friends, I learned one of the defining features of Friends' successful original run was its ratings/viewership consistency; according to Nielsen Media Research, Friends never ranked below #5 among primetime television shows after its first season, while reaching #1 during Season 8 (2001-2002; Link: https://en.wikipedia.org/wiki/Friends).

As such, I began my analysis by examining the overarching season level viewership data for the series. Since TV shows such as Friends could have "outlier" viewership numbers for 1 or 2 episodes on account of a season/series premiere or finale, specials such as airing directly after the Super Bowl in 1996, and even flashback episodes (which largely re-hash storylines from prior episodes), I decided to use the median to perform this analysis, as the median can be more robust than the mean to outliers.

I calculated the median episode by viewership (in millions of viewers) for each season, which is shown in the table below. I also calculated the percentage change in viewership from season to season (by its median viewership episode).

```r
season.median.viewership.summary = friends.database %>%
  group_by(Season) %>%
  summarise(Median.Viewership = median(Viewers))

season.viewership.percentage.change = rep(NA, nrow(season.median.viewership.summary))

# Calculate Percentage Change in Viewership Between "Median" Episodes from Sequential Seasons
for(i in 2:nrow(season.median.viewership.summary)){
  season.viewership.percentage.change[i] = (season.median.viewership.summary$Median.Viewership[i] -
    season.median.viewership.summary$Median.Viewership[i - 1]) /
    season.median.viewership.summary$Median.Viewership[i] * 100
}

season.median.viewership.summary = cbind.data.frame(season.median.viewership.summary,
    season.viewership.percentage.change)

season.median.viewership.summary = rename(season.median.viewership.summary,
    Viewership.Percentage.Change = season.viewership.percentage.change)

season.median.viewership.summary$Viewership.Percentage.Change = round(season.median.viewership.summary$V
    digits = 2)

print(season.median.viewership.summary)
```

```
##    Season Median.Viewership Viewership.Percentage.Change
## 1       1            24.400                           NA
## 2       2            30.200                        19.21
```

```
## 3        3              26.100                        -15.71
## 4        4              24.330                         -7.27
## 5        5              24.870                          2.17
## 6        6              22.330                        -11.37
## 7        7              22.320                         -0.04
## 8        8              25.970                         14.05
## 9        9              24.065                         -7.92
## 10      10              23.545                         -2.21
```

The table above confirms the consistency in the Nielsen's rankings data, as Friends was able to sustain a median 22+ million viewers each season for 10 years; furthermore, Friends only had one 15% decline in median viewership between seasons (which occured between Seasons 2 and 3). This fact highlights Friends' ability to retain its audience.

```r
season.imdb.rating.summary = friends.database %>%
  group_by(Season) %>%
  summarise(Median.IMDB.Rating = median(IMDB_Rating),
            Twenty.Fifth.Percentile = quantile(IMDB_Rating, probs = 0.25),
            Seventy.Fifth.Percentile = quantile(IMDB_Rating, probs = 0.75)) %>%
  select(Season, Twenty.Fifth.Percentile, Median.IMDB.Rating, Seventy.Fifth.Percentile)

print(season.imdb.rating.summary)
```

```
## # A tibble: 10 x 4
##    Season Twenty.Fifth.Percentile Median.IMDB.Rating Seventy.Fifth.Percentile
##    <fct>                    <dbl>              <dbl>                    <dbl>
##  1 1                         8.1                8.2                      8.5
##  2 2                         8.1                8.35                     8.6
##  3 3                         8.1                8.3                      8.6
##  4 4                         8.2                8.5                      8.7
##  5 5                         8.28               8.55                     8.93
##  6 6                         8.2                8.5                      8.6
##  7 7                         8.28               8.4                      8.6
##  8 8                         8.17               8.3                      8.7
##  9 9                         8.1                8.2                      8.52
## 10 10                        8.35               8.6                      8.9
```

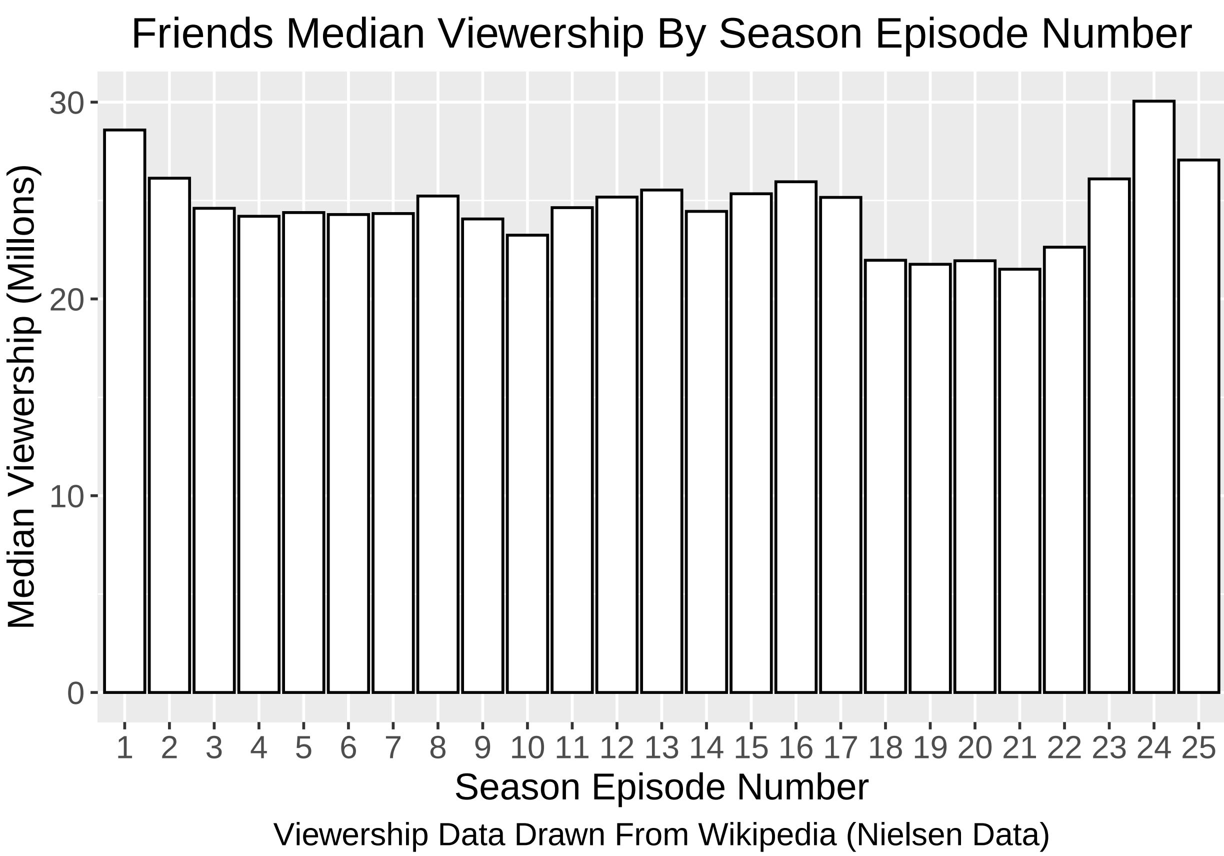
To see whether such consistency extended to viewer reception as well, I calculated the interquartile range (25th percentile, 50th percentile or median, and 75th percentile) of IMDB scores by season. Although the IMDB data is not retrospective like the Nielsen data (from the orginial air dates), this data gives an idea of how audiences (both current and past) might have reacted to each episode.

The interquartile IMDB data by season is shown above, where there is basically no difference in IMDB score by season, with all seasons posting strong 8+ scores out of 10.

Taking this analysis one step further, I wanted to investigate how well Friends was able to retain its audience/viewership acclaim not just between seasons, but also intraseason as well (as a 24 episode season spanned nearly 8 months of television.) As such, I calculated the median viewership by season episode number across all 10 seasons, as well as the percentage change between sequential medians.

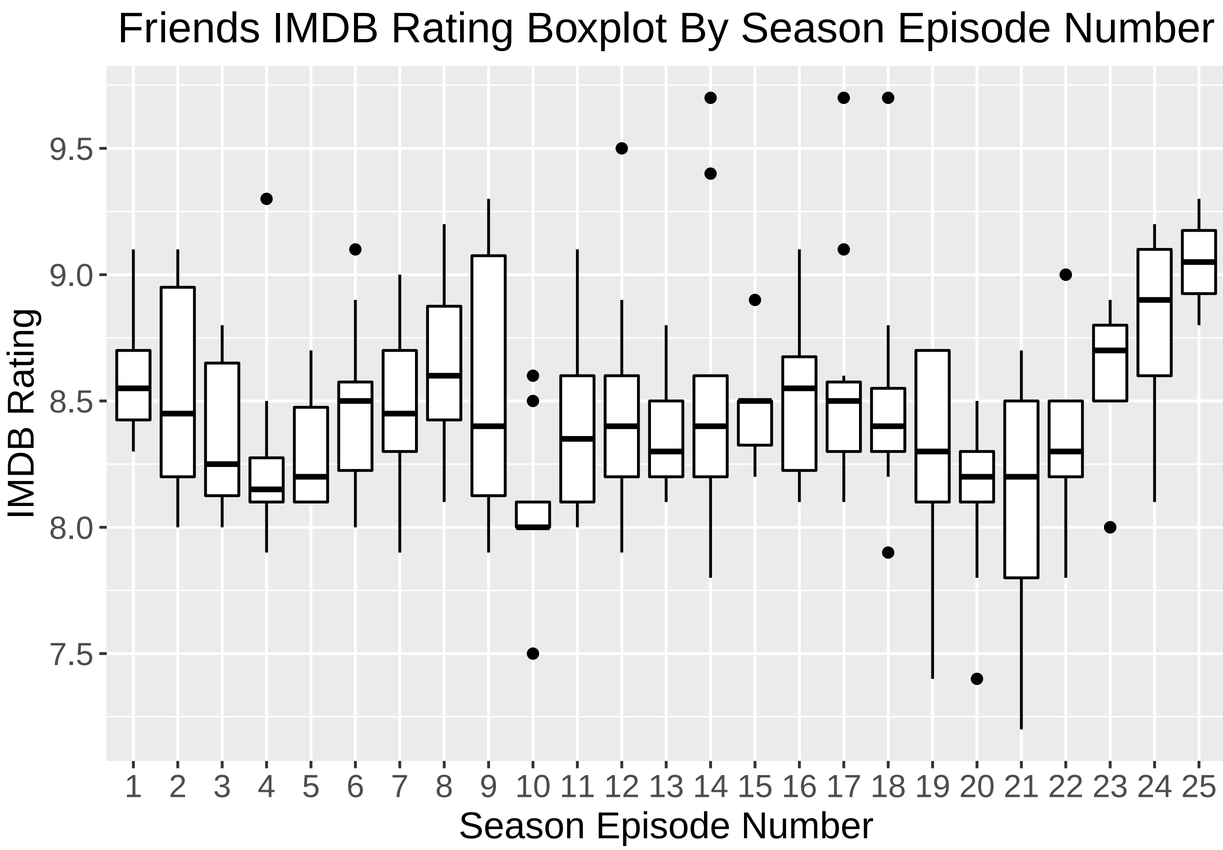
```r
episode.median.viewership.summary = friends.database %>%
  group_by(Episode) %>%
  summarise(Median.Viewership = median(Viewers))

episode.viewership.percentage.change = rep(NA, nrow(episode.median.viewership.summary))

# Calculate Percentage Change in Viewership Between Episodes
```

```
for(i in 2:nrow(episode.median.viewership.summary)){
  episode.viewership.percentage.change[i] = (episode.median.viewership.summary$Median.Viewership[i] -
    episode.median.viewership.summary$Median.Viewership[i - 1]) /
    episode.median.viewership.summary$Median.Viewership[i] * 100
}

episode.median.viewership.summary = cbind.data.frame(episode.median.viewership.summary,
    episode.viewership.percentage.change)

episode.median.viewership.summary = rename(episode.median.viewership.summary,
    Viewership.Percentage.Change = episode.viewership.percentage.change)

episode.median.viewership.summary$Viewership.Percentage.Change = round(episode.median.viewership.summar
                                                                     digits = 2)

median.viewership.by.season.episode.number.barplot = ggplot(episode.median.viewership.summary,
    aes(x = Episode, y = Median.Viewership)) + geom_bar(stat = "identity", fill = "white",
        color = "black") + theme(axis.text = element_text(size = 12),
        axis.title = element_text(size = 14), plot.title = element_text(hjust = 0.5, size = 16),
        plot.caption = element_text(size = 12, hjust = 0.5)) +
    labs(title = "Friends Median Viewership By Season Episode Number",
       x = "Season Episode Number", y = "Median Viewership (Millons)",
       caption = "Viewership Data Drawn From Wikipedia (Nielsen Data)")

print(median.viewership.by.season.episode.number.barplot)
```



As shown in the barplot above, Friends achieved noticable consistency, irrespective of episode number within a season. While the season premiere (Episode #1) and frequent season finale (Episode #24) had slightly higher median viewership, all episode numbers had a median viewership total in excess of 20 million. Furthermore, there was only one noteworthy drop in viewers, between Episodes 17 and 18, namely as the show would generally be entering its final quarter of episodes for a season (but prior to the finale).

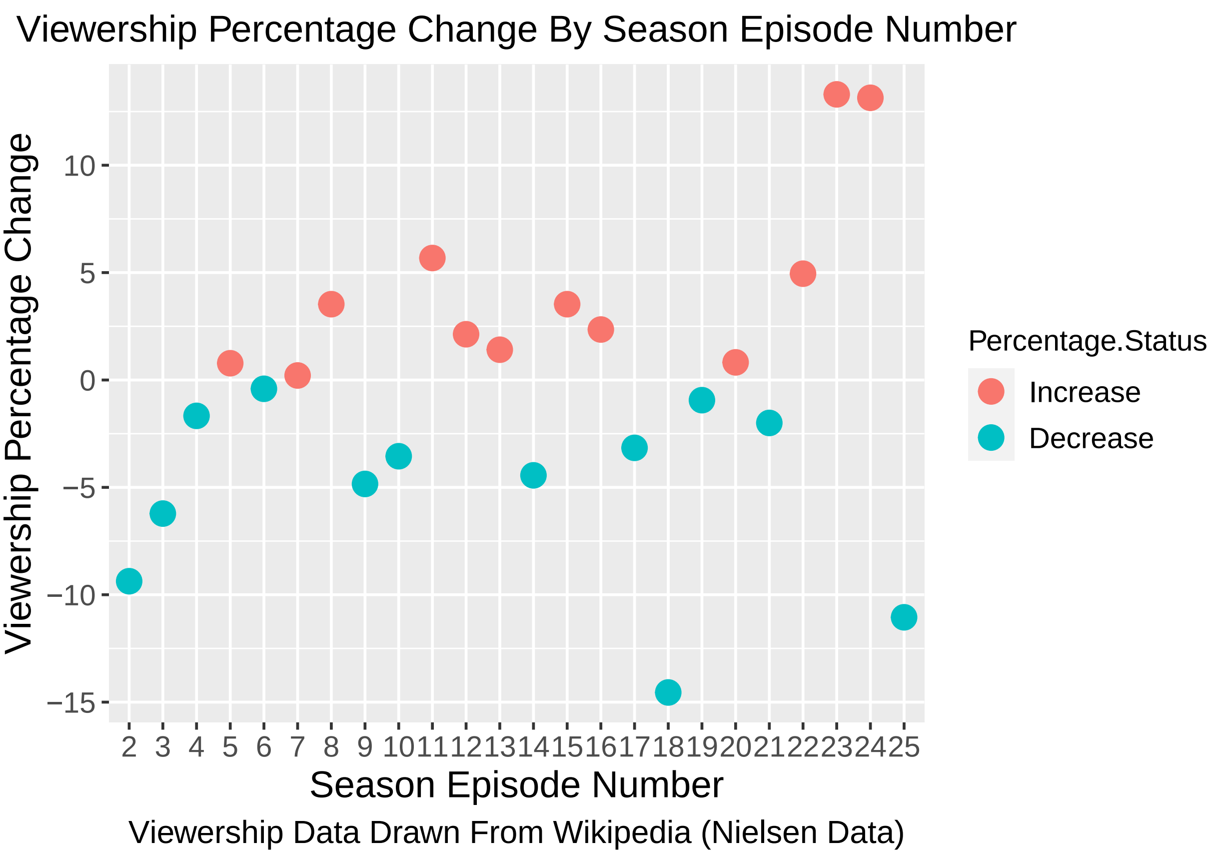
```
IMDB.rating.by.season.episode.number.boxplot = ggplot(friends.database,
    aes(x = Episode, y = IMDB_Rating)) + geom_boxplot(fill = "white",
        color = "black") + theme(axis.text = element_text(size = 12),
        axis.title = element_text(size = 14),
        plot.title = element_text(hjust = 0.5, size = 16),
        plot.caption = element_text(size = 12, hjust = 0.5)) +
    labs(title = "Friends IMDB Rating Boxplot By Season Episode Number",
      x = "Season Episode Number", y = "IMDB Rating") +
   scale_y_continuous(breaks = seq(6, 10, 0.5))

print(IMDB.rating.by.season.episode.number.boxplot)
```



Similarly, the boxplot of IMDB data by season episode number shows season finales (Episodes 23-25) were the most critically acclaimed among audiences, while Episodes 19-22 had a slight drop in median ratings; this is consisent with the viewership finding for these episodes, and suggests that a possible explanation for the slight drop in viewership was the fact that those episodes (leading into the finale) were not as well received by audiences.

```
episode.median.viewership.summary = mutate(episode.median.viewership.summary,
  Percentage.Status =  ifelse(Viewership.Percentage.Change >= 0, "Increase", "Decrease"))

episode.median.viewership.summary$Percentage.Status = factor(episode.median.viewership.summary$Percentag
    levels = c("Increase", "Decrease"))
```

```r
median.viewership.by.season.episode.number.dotplot = ggplot(filter(episode.median.viewership.summary,
  !is.na(Percentage.Status)), aes(x = Episode, y = Viewership.Percentage.Change,
    colour = Percentage.Status)) + geom_point(size = 4) +
  labs(title = "Viewership Percentage Change By Season Episode Number",
      x = "Season Episode Number", y = "Viewership Percentage Change",
      caption = "Viewership Data Drawn From Wikipedia (Nielsen Data)") +
  theme(axis.text = element_text(size = 11),
        axis.title = element_text(size = 14),
        plot.title = element_text(hjust = 0.5, size = 14),
        legend.text = element_text(size = 11),
        plot.caption = element_text(size = 12, hjust = 0.5))

print(median.viewership.by.season.episode.number.dotplot)
```



The dot plot above shows the percentage change between median viewership by season episode numbers. From this chart, we see the increase in viewers heading into a season finale (Episodes 23 and 24), while there was only drop of greater than ~10% between Episode 17 and 18 (described above).

In seeing the lower viewership numbers, in the aggregate, for Episodes 18-22 of a season, I wanted to next examine whether this decline could also possibly be attributed to a seasonal effect (vs audience reception from the IMDB data).

As such, I calculated the median viewership total by month across all 10 seasons of Friends in the following table.

```r
# Viewership for Median Episode by Month
month.median.viewership.summary = friends.database %>%
  group_by(Month) %>%
```

```r
  summarise(Median.Viewership = median(Viewers))


print(month.median.viewership.summary)
```

```
## # A tibble: 9 x 2
##   Month Median.Viewership
##   <ord>             <dbl>
## 1 Jan                25.8
## 2 Feb                25.1
## 3 Mar                23.6
## 4 Apr                21.9
## 5 May                25.9
## 6 Sep                27.7
## 7 Oct                24.4
## 8 Nov                24.4
## 9 Dec                23.2
```

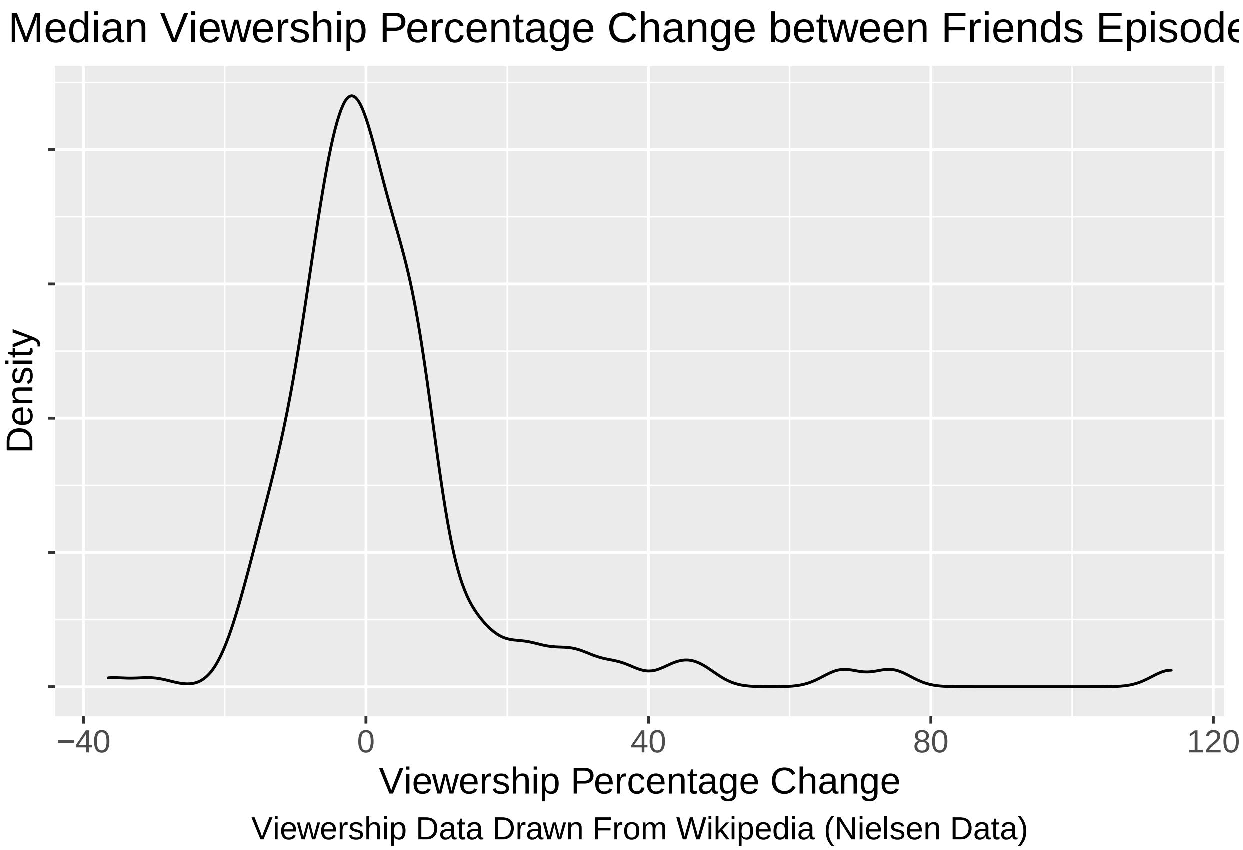
This table shows the highest median viewership in September (the month of the season premiere), as well as May (the month of season finale) and January (the start of the calendar year). In turn, while there is a slight dip in median viewership in April (heading into the May finale), the difference is rather small. As such, this result further suggests that the fourth quarter dip in ratings could be due to a non-seasonal effect (such as show mechanics/audience reception), rather than a seasonal one.

```r
viewership.percentage.change = rep(NA, nrow(friends.database))

# Calculate Percentage Change in Viewership Between Episodes
for(i in 2:nrow(friends.database)){
  date.one = friends.database$Air.Date[i - 1]
  date.two = friends.database$Air.Date[i]
  if(year(date.one) == year(date.two) & month(date.one) == month(date.two) &
     day(date.one) == day(date.two)){
    viewership.percentage.change[i] = ((friends.database$Viewers[i] -
      friends.database$Viewers[i - 2]) / friends.database$Viewers[i - 2] * 100)
  }else{
    viewership.percentage.change[i] = ((friends.database$Viewers[i] -
      friends.database$Viewers[i - 1]) / friends.database$Viewers[i - 1] * 100)
  }
}

friends.database = cbind.data.frame(friends.database, viewership.percentage.change)

friends.database$viewership.percentage.change = round(friends.database$viewership.percentage.change,
                                                      digits = 2)
```

I next calculated the percentage change in viewership for all 236 episodes, and plotted these deltas in a density plot.

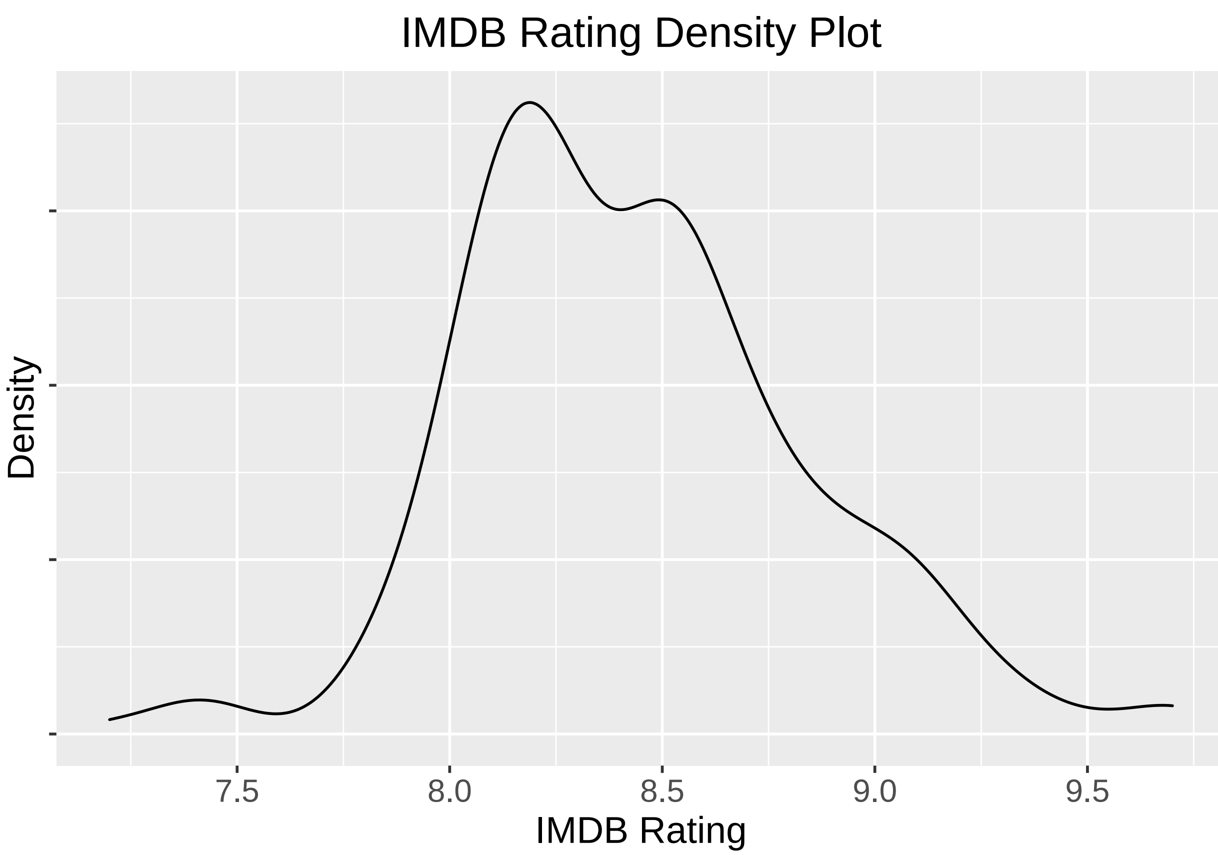
```r
viewership.change.density.plot = ggplot(friends.database,
  aes(x = viewership.percentage.change)) +
  geom_density() + theme(axis.text = element_text(size = 12),
       axis.title = element_text(size = 14),
       plot.title = element_text(hjust = 0.5, size = 16),
       axis.text.y = element_blank(),
       plot.caption = element_text(size = 12, hjust = 0.5)) +
```

```
       labs(title = "Median Viewership Percentage Change between Friends Episodes",
       x = "Viewership Percentage Change", y = "Density",
       caption = "Viewership Data Drawn From Wikipedia (Nielsen Data)")

print(viewership.change.density.plot)
```



As shown above, the vast majority of deltas from episode to episode are centered around 0 (confirming prior results), with a very small portion of the distribution showing a greater than 20% decline episode to episode. The distribution is also right-skewed, indicating that Friends did have more significant increases in viewership episode to episode, rather than declines.

```
IMDB.rating.density.plot = ggplot(friends.database,
  aes(x = IMDB_Rating)) +
  geom_density() + theme(axis.text = element_text(size = 12),
       axis.title = element_text(size = 14),
       plot.title = element_text(hjust = 0.5, size = 16),
       axis.text.y = element_blank(),
       plot.caption = element_text(size = 12, hjust = 0.5)) +
       labs(title = "IMDB Rating Density Plot",
       x = "IMDB Rating", y = "Density") + scale_x_continuous(breaks = seq(6, 10, 0.5))

print(IMDB.rating.density.plot)
```



I also calculated a density plot of IMDB scores across all 236 episodes. This density plot shows the modal IMDB Rating for an episode as being ~8.2; this density curve is also right skewed from the mode, indicating a larger number of episodes with even higher acclaim from audiences (IMDB Ratings of 9+) than less critically successful shows.

```
friends.database$Writer_1 = as.character(friends.database$Writer_1)
friends.database$Writer_2 = as.character(friends.database$Writer_2)
friends.database$Writer_3 = as.character(friends.database$Writer_3)
friends.database$Writer_4 = as.character(friends.database$Writer_4)

# Remove Whitespace from Friends Writer Name strings
friends.database$Writer_1 = gsub("(^\\s+)|(\\s+$)", "", friends.database$Writer_1)
friends.database$Writer_2 = gsub("(^\\s+)|(\\s+$)", "", friends.database$Writer_2)
friends.database$Writer_3 = gsub("(^\\s+)|(\\s+$)", "", friends.database$Writer_3)
friends.database$Writer_4 = gsub("(^\\s+)|(\\s+$)", "", friends.database$Writer_4)

# Standardize Writing Credit Names
friends.database$Writer_1 = gsub("Sherry Bilsing-Graham", "Sherry Bilsing",
                                 friends.database$Writer_1)
friends.database$Writer_2 = gsub("Sherry Bilsing-Graham", "Sherry Bilsing",
                                 friends.database$Writer_2)
friends.database$Writer_3 = gsub("Sherry Bilsing-Graham", "Sherry Bilsing",
                                 friends.database$Writer_3)
friends.database$Writer_4 = gsub("Sherry Bilsing-Graham", "Sherry Bilsing",
                                 friends.database$Writer_4)

friends.writers = friends.database %>%
```

```r
  select(Writer_1, Writer_2, Writer_3, Writer_4) %>%
  gather(Writer_Position, Writer) %>%
  select(Writer) %>%
  filter(Writer != "") %>%
  unique()
```

After the analyses above, I next looked at the possible effect of director selection on Friends viewership totals and IMDB Ratings.

```r
# Number of Episodes and Median Viewer by Director
friends.directors.summary.statistics = friends.database %>%
  group_by(Director) %>%
  summarise(Episode.Count = n(), Median.Viewers = median(Viewers),
            Median.IMDB.Rating = median(IMDB_Rating)) %>%
  arrange(desc(Median.Viewers)) %>%
  filter(Episode.Count >= 10)

print(friends.directors.summary.statistics)
```

```
## # A tibble: 8 x 4
##   Director        Episode.Count Median.Viewers Median.IMDB.Rating
##   <fct>                   <int>          <dbl>              <dbl>
## 1 Michael Lembeck            24           27.8               8.5
## 2 Kevin S. Bright            53           25.9               8.7
## 3 Gail Mancuso               14           25.8               8.35
## 4 David Schwimmer            10           24.1               8.5
## 5 Peter Bonerz               12           23.9               8.2
## 6 James Burrows              15           23.8               8.3
## 7 Ben Weiss                  10           23.0               8.25
## 8 Gary Halvorson             54           22.6               8.3
```

As shown in the table above, 8 directors directed at least 10 solo episodes over the 236 episode show (Kevin S. Bright and Gary Halvorson also directed one episode jointly). Of these eight directors, Michael Lembeck achieved the highest median viewership in his 24 episodes, followed by Kevin Bright (who along with Marta Kaufmann and David Crane produced Friends). On the other hand, Gary Halvorson, who directed the most episodes of 55 (54 solo, 1 joint), interestingly had the lowest median viewership. In addition, while the margins between median IMDB ratings are very small, there is appears to be at least a bit of a correlation between median viewership and median IMDB rating (i.e. audience reception moving lower with viewership and vice-versa by director).

I next performed a similar analysis for writers and writers' credits.

```r
# Number of Episodes and Median Viewer by Writer (Story and Teleplay Credits)
friends.writers.summary.statistics = friends.database %>%
  select(Writer_1, Writer_2, Writer_3, Writer_4, Viewers, IMDB_Rating) %>%
  gather(Writer_Position, Writer, -Viewers, -IMDB_Rating) %>%
  select(Writer, Viewers, IMDB_Rating) %>%
  filter(Writer != "") %>%
  group_by(Writer) %>%
  summarise(Episode.Count = n(), Median.Viewers = median(Viewers),
            Median.IMDB.Rating = median(IMDB_Rating)) %>%
  arrange(desc(Median.Viewers)) %>%
  filter(Episode.Count > 10)

print(friends.writers.summary.statistics)
```

```
## # A tibble: 16 x 4
##    Writer               Episode.Count Median.Viewers Median.IMDB.Rating
##    <chr>                        <int>          <dbl>              <dbl>
##  1 Alexa Junge                     12           27.2               8.35
##  2 David Crane                     21           26.8               8.6
##  3 Marta Kauffman                  21           26.8               8.6
##  4 Adam Chase                      16           26.1               8.35
##  5 Gregory S. Malins               17           25.1               8.5
##  6 Wil Calhoun                     11           25.1               8.2
##  7 Michael Curtis                  12           25.0               8.35
##  8 Andrew Reich                    25           23.9               8.5
##  9 Ted Cohen                       25           23.9               8.5
## 10 Scott Silveri                   23           23.7               8.5
## 11 Seth Kurland                    13           23.7               8.3
## 12 Shana Goldberg-Meehan           22           23.5               8.3
## 13 Ellen Plummer                   17           22.7               8.4
## 14 Sherry Bilsing                  17           22.7               8.4
## 15 Brian Buckner                   15           22.2               8.3
## 16 Sebastian Jones                 15           22.2               8.2
```

Of the sixteen writers with at least 10 writing (story + teleplay) credits during the 10-year show, Alexa Junge had the highest median viewership total for her 12 episodes (followed by show creators David Crane and Marta Kauffman). While there was a slight dip when median viewership was broken down by director, there was no such dip by writer. Furthermore, there is no evidence of correlation between median viewership and median IMDB rating by writer. Of note, show creators David Crane and Marta Kauffman had the highest median IMDB scores for their episodes from audiences.

Finally, I wanted to examine all of the variables considered above together for their combined effect on viewership (a quantitative response variable). Since the goal of such analysis is inference rather than prediction, we can utilize a low-variance statistical learning/machine learning model like linear regression. As such, I created a linear regression model, regressing episode viewership across the 236 Friends episodes with Season Number, Episode Number, Director, the Month the episode originally aired, and whether series co-creators David Crane and Marta Kauffman served as writers (as the analysis above indicated that they had written for among the most watched episodes in the series, and critically acclaimed episodes by audiences.)

```r
# Features and Response Variable for Linear Regression Model
friends.linear.regression.features = friends.database %>%
  # Boolean Indicating Writing Credits for Marta Kauffman and David Crane
  mutate(Crane.Kauffman.Written.Episode = ifelse(Writer_1 == "David Crane",
         "Yes", "No")) %>%
  select(-Month) %>%
  mutate(Month = month(Air.Date)) %>%
  select(Season, Episode, Director, Viewers, Crane.Kauffman.Written.Episode, Month)

friends.linear.regression.features$Month = factor(friends.linear.regression.features$Month,
                                                  levels = seq(1, 12))

# Regressing Episode Viewers on list of Features
friends.linear.regression.model = lm(Viewers ~.,
      data = friends.linear.regression.features)

# Linear Regression Model Output
summary(friends.linear.regression.model)
```

```
##
## Call:
```

```
## lm(formula = Viewers ~ ., data = friends.linear.regression.features)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -7.851 -2.026  0.000  1.580 17.358
##
## Coefficients:
##                                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)                               37.5086     6.2975   5.956 1.51e-08
## Season2                                    3.8166     1.6888   2.260  0.02513
## Season3                                    0.7177     1.5620   0.459  0.64651
## Season4                                   -2.1567     1.5662  -1.377  0.17037
## Season5                                   -2.1074     1.7521  -1.203  0.23080
## Season6                                   -3.4133     1.7292  -1.974  0.05006
## Season7                                   -5.0589     1.8134  -2.790  0.00590
## Season8                                   -0.4351     1.7945  -0.242  0.80872
## Season9                                   -3.1644     1.8280  -1.731  0.08531
## Season10                                  -2.5604     2.0823  -1.230  0.22060
## Episode2                                  -2.9102     2.1847  -1.332  0.18467
## Episode3                                  -5.1503     2.7881  -1.847  0.06651
## Episode4                                  -5.2775     2.8233  -1.869  0.06336
## Episode5                                  -4.8755     2.9724  -1.640  0.10286
## Episode6                                 -11.2412     4.0698  -2.762  0.00639
## Episode7                                 -11.9523     4.3119  -2.772  0.00621
## Episode8                                 -12.9662     4.3382  -2.989  0.00323
## Episode9                                 -13.1095     4.6313  -2.831  0.00522
## Episode10                                -13.4492     5.1793  -2.597  0.01026
## Episode11                                -13.4320     5.5294  -2.429  0.01621
## Episode12                                -12.3093     5.6288  -2.187  0.03016
## Episode13                                 -9.8224     5.7777  -1.700  0.09100
## Episode14                                -11.3510     6.0424  -1.879  0.06207
## Episode15                                -11.4328     6.1747  -1.852  0.06588
## Episode16                                -13.2855     6.2019  -2.142  0.03365
## Episode17                                -11.0810     6.3731  -1.739  0.08395
## Episode18                                -11.1390     6.6372  -1.678  0.09519
## Episode19                                -15.7349     6.7342  -2.337  0.02066
## Episode20                                -18.0045     6.9150  -2.604  0.01006
## Episode21                                -22.0812     7.0719  -3.122  0.00212
## Episode22                                -24.0418     7.2382  -3.321  0.00110
## Episode23                                -22.4748     7.2836  -3.086  0.00238
## Episode24                                -22.0924     7.2636  -3.042  0.00274
## Episode25                                -18.7399     8.0436  -2.330  0.02103
## DirectorAndrew Tsao                        7.0662     5.4914   1.287  0.19998
## DirectorArlene Sanford                     2.9208     5.9212   0.493  0.62247
## DirectorBen Weiss                          3.9783     3.7044   1.074  0.28441
## DirectorDana DeVally Piazza                3.8388     4.1336   0.929  0.35441
## DirectorDavid Schwimmer                    5.1972     3.7447   1.388  0.16705
## DirectorDavid Steinberg                    1.7401     5.4785   0.318  0.75118
## DirectorEllen Gittelsohn                   2.8385     5.5054   0.516  0.60684
## DirectorGail Mancuso                       4.6440     3.4943   1.329  0.18568
## DirectorGary Halvorson                     2.5934     3.4871   0.744  0.45811
## DirectorJames Burrows                      1.6359     3.2122   0.509  0.61123
## DirectorJoe Regalbuto                      5.0508     5.5394   0.912  0.36321
## DirectorKevin S. Bright                    5.9875     3.4670   1.727  0.08604
```

```
## DirectorKevin S. Bright & Gary Halvorson   0.2479     5.5609   0.045  0.96450
## DirectorMary Kay Place                     6.1799     5.6396   1.096  0.27475
## DirectorMichael Lembeck                    6.5264     3.4369   1.899  0.05932
## DirectorPamela Fryman                     -3.8578     4.5919  -0.840  0.40205
## DirectorPaul Lazarus                      -1.1992     5.1371  -0.233  0.81571
## DirectorPeter Bonerz                       7.2096     3.4830   2.070  0.04002
## DirectorRobby Benson                       3.2356     3.6751   0.880  0.37992
## DirectorRoger Christiansen                 4.3498     4.6694   0.932  0.35293
## DirectorSam Simon                          2.3886     5.4987   0.434  0.66457
## DirectorSheldon Epps                       7.3810     4.3226   1.708  0.08960
## DirectorShelley Jensen                     4.3250     3.7892   1.141  0.25536
## DirectorStephen Prime                      2.7747     5.5205   0.503  0.61591
## DirectorSteve Zuckerman                    2.5133     4.5862   0.548  0.58442
## DirectorTerry Hughes                       3.9445     3.9566   0.997  0.32024
## DirectorThomas Schlamme                    4.0155     4.7996   0.837  0.40401
## DirectorTodd Holland                      -1.3586     5.6139  -0.242  0.80908
## Crane.Kauffman.Written.EpisodeYes          2.0811     1.1952   1.741  0.08351
## Month2                                    -3.1619     2.0482  -1.544  0.12457
## Month3                                    -3.6938     2.9928  -1.234  0.21887
## Month4                                    -1.0236     3.3676  -0.304  0.76154
## Month5                                     8.5743     3.8278   2.240  0.02642
## Month9                                   -13.8439     5.4590  -2.536  0.01214
## Month10                                  -10.9881     4.7964  -2.291  0.02323
## Month11                                   -3.7200     3.5437  -1.050  0.29537
## Month12                                   -3.7253     2.5475  -1.462  0.14555
##
## (Intercept)                               ***
## Season2                                   *
## Season3
## Season4
## Season5
## Season6                                   .
## Season7                                   **
## Season8
## Season9                                   .
## Season10
## Episode2
## Episode3                                  .
## Episode4                                  .
## Episode5
## Episode6                                  **
## Episode7                                  **
## Episode8                                  **
## Episode9                                  **
## Episode10                                 *
## Episode11                                 *
## Episode12                                 *
## Episode13                                 .
## Episode14                                 .
## Episode15                                 .
## Episode16                                 *
## Episode17                                 .
## Episode18                                 .
## Episode19                                 *
```

```
## Episode20                                *
## Episode21                                **
## Episode22                                **
## Episode23                                **
## Episode24                                **
## Episode25                                *
## DirectorAndrew Tsao
## DirectorArlene Sanford
## DirectorBen Weiss
## DirectorDana DeVally Piazza
## DirectorDavid Schwimmer
## DirectorDavid Steinberg
## DirectorEllen Gittelsohn
## DirectorGail Mancuso
## DirectorGary Halvorson
## DirectorJames Burrows
## DirectorJoe Regalbuto
## DirectorKevin S. Bright                  .
## DirectorKevin S. Bright & Gary Halvorson
## DirectorMary Kay Place
## DirectorMichael Lembeck                  .
## DirectorPamela Fryman
## DirectorPaul Lazarus
## DirectorPeter Bonerz                     *
## DirectorRobby Benson
## DirectorRoger Christiansen
## DirectorSam Simon
## DirectorSheldon Epps                     .
## DirectorShelley Jensen
## DirectorStephen Prime
## DirectorSteve Zuckerman
## DirectorTerry Hughes
## DirectorThomas Schlamme
## DirectorTodd Holland
## Crane.Kauffman.Written.EpisodeYes        .
## Month2
## Month3
## Month4
## Month5                                   *
## Month9                                   *
## Month10                                  *
## Month11
## Month12
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.96 on 165 degrees of freedom
## Multiple R-squared:  0.5945, Adjusted R-squared:  0.4225
## F-statistic: 3.456 on 70 and 165 DF,  p-value: 3.779e-11
```

The linear regression model summary, as shown above, indicates a statistically significant positive Season 2 effect (at a standard p-value of 0.05) as compared to Season 1 on viewership, while there is a statistically significant negative effect for Season 7 as compared to Season 1 (with no other seasons having statistically significant p-values). Taken together, the results could indicate that positive buzz from a successful first season, coupled with Friends' prominent feature in the post-Super Bowl timeslot, as well as the long sought-after Ross Rachel relationship by fans, all resulted in a significant positive Season 2 effect on viewership when compared to Season 1. Likewise, Season 7 had a significant negative effect; nevertheless, Season 6's coefficient was not statistically signficant, so one long hypothesized explanation for the Season 7 dip (i.e. Season 7's focus on Monica and Chandler's relationship) may not hold. Instead, Season 7 could possibly be explained as an inverse of Season 2, namely a seeming lack of focus on Ross and Rachel's relationship (which sits between Ross and Rachel's divorce in Season 6, and Ross and Rachel's pregnancy in Season 8).

In turn, episodes in the second and fourth quarters of the TV season (Episodes 6-12 and Episodes 18-24) had statistically significant negative coefficients/effects on viewership as compared to season premieres. Furthermore, while May, September, and October's effects on viewership as compared to January were positive, negative, negative, respectively, November, December, March and April's coefficients were not statistically significant. Taken together, this suggests that show mechanics/non-seasonal effects could better explain any dip in viewership, rather than a seasonal effect (such as any potential negative effect of the holidays Thanksgiving or Christmas on viewership). The second and fourth quarters for Friends could be when the writers and producers had to work on character/storyline development to set up a suprise in the middle of the season (such as Ross and Rachel's initial relationship in Season 2 Episode 14) or the end of the season (such as Ross's wedding to Emily, Chandler's proposal and wedding to Monica etc.).

Finally, with the exception of a positive coefficient/effect for Director Peter Bonerz, the remaining director and writer credit coefficients were not statistically significant. However, one potential limitation of the model is that it represents only ~42% in the variation in viewership (adjusted R^2 value). More specifically, due to a limited number of features available in the Wikipedia dataset (which did not include features that reflect acting performances, elements from the scripts themselves or thematic focuses etc.), there is a lot of information toward understanding Friends viewership numbers that cannot be explained by the features above. Furthermore, director and writers credits do not capture the behind-the-scenes effect that a team of writers, producers and creators had on the show's success. At the same time, the model and analysis can be useful in possibly ruling out possible explanations, or coming up with new hypotheses that could be further explored in further analysis.

IV: Summary

- Friends' viewership and audience reception (IMDB rating) was largely consistent both inter-season and intra-season
- There was a significant bounce in viewership in Season 2 after Season 1 (that cannot be explained by the post Super Bowl episodes alone)
- Viewership and audience reception was softer in the fourth quarter of the TV season heading into the season finale (which possibly could be explained by necessary character/storyline development ahead of a surprising/impactful finale)
- Director/Writing credits did not appear to impact ratings; a better possible explanation of Friends' ratings strength is the team of producers, writers and creators beyond the credits
- Cast performances undoubtedly had a significant impact on Friends' sustained success (and cannot be quantified in this dataset).