

video_game_data

Oscar Monroy

2/22/2022

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 3.6.3
```

```
##
```

```
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
```

```
##
```

```
## filter, lag
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
## intersect, setdiff, setequal, union
```

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.6.3
```

```
library(tidyr)
```

```
## Warning: package 'tidyr' was built under R version 3.6.3
```

```
vgs <- read.csv("game_sales_data.csv")
```

```
summary(vgs) # Seems mostly normal except for the NA's
```

```
##      Rank      Name      Platform
## Min.   :    1  FIFA 14      :    9  DS      :2251
## 1st Qu.: 4899  Madden NFL 07 :    9  PS2     :2214
## Median : 9798  Madden NFL 08 :    9  PC      :1913
## Mean   : 9799  Minecraft      :    9  PS3     :1368
## 3rd Qu.:14698  Ratatouille    :    9  Wii     :1352
## Max.   :19598  Angry Birds: Star Wars:    8  PSP     :1313
##              (Other)      :19547  (Other):9189
##      Publisher      Developer      Critic_Score
## Activision      : 1024  Unknown      : 489  Min.   : 0.800
## Ubisoft         : 950  Konami      : 402  1st Qu.: 6.100
## Electronic Arts: 829  Capcom      : 350  Median : 7.300
```

```
## Konami      : 768 EA Canada      : 286 Mean    : 7.035
## Nintendo    : 765 Bandai Namco Games: 226 3rd Qu.: 8.200
## THQ         : 717 EA Tiburon     : 195 Max.    :10.000
## (Other)     :14547 (Other)       :17652 NA's    :9631
## User_Score  Total_Shipped      Year
## Min.       : 1.000 Min.       : 0.0100 Min.       :1977
## 1st Qu.: 6.300 1st Qu.: 0.0500 1st Qu.:2004
## Median : 7.200 Median : 0.1600 Median :2008
## Mean      : 6.995 Mean      : 0.5511 Mean      :2008
## 3rd Qu.: 8.000 3rd Qu.: 0.4600 3rd Qu.:2012
## Max.      :10.000 Max.      :82.9000 Max.      :2020
## NA's      :17377
```

```
sum(duplicated(vgs)) # Nor duplicated rows, luckily
```

```
## [1] 0
```

```
head(vgs) # These 6 games are the highest selling (physical) games of all time
```

```
## Rank Name Platform Publisher
## 1 1 Wii Sports Wii Nintendo
## 2 2 Super Mario Bros. NES Nintendo
## 3 3 Counter-Strike: Global Offensive PC Valve
## 4 4 Mario Kart Wii Wii Nintendo
## 5 5 PLAYERUNKNOWN'S BATTLEGROUNDS PC PUBG Corporation
## 6 6 Minecraft PC Mojang
## Developer Critic_Score User_Score Total_Shipped Year
## 1 Nintendo EAD 7.7 8.0 82.90 2006
## 2 Nintendo EAD 10.0 8.2 40.24 1985
## 3 Valve Corporation 8.0 7.5 40.00 2012
## 4 Nintendo EAD 8.2 9.1 37.32 2008
## 5 PUBG Corporation 8.6 4.7 36.60 2017
## 6 Mojang AB 10.0 7.8 33.15 2010
```

Visualization of the Average Review Score and Shipments Throughout the Years

```
vgs_year_s <- vgs %>%
  select(Year, Critic_Score, User_Score, Total_Shipped) %>%
  group_by(Year) %>%
  summarize(
    avg_critic_score = mean(Critic_Score, na.rm = T),
    avg_user_score = mean(User_Score, na.rm = T),
    avg_total_shipped = mean(Total_Shipped, na.rm = T)
  )
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

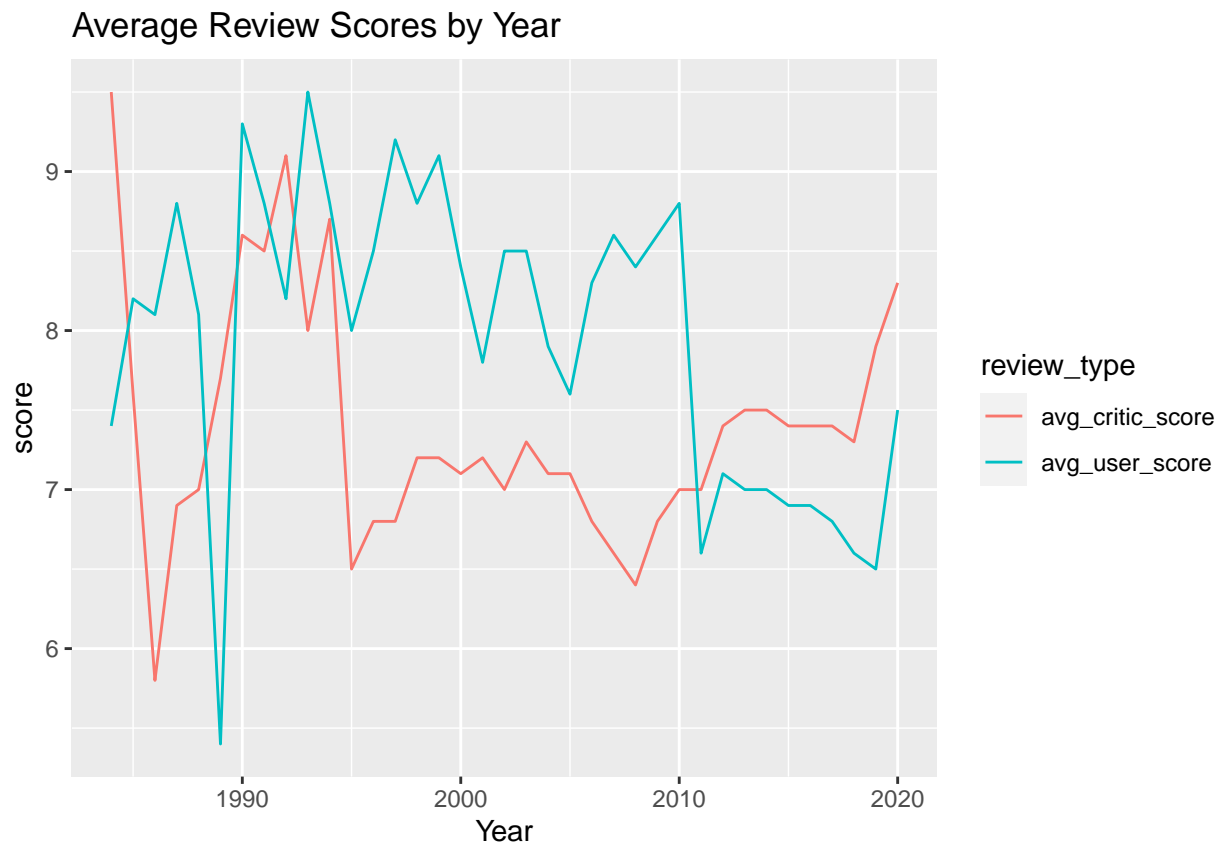
```

vgs_year_s[, c(2, 3)] <- round(vgs_year_s[, c(2, 3)], 1)
vgs_year_s1 <- vgs_year_s[-c(1:7), ]

review_scores <- vgs_year_s1 %>%
  pivot_longer(cols = "avg_critic_score":"avg_user_score",
               names_to = "review_type", values_to = "score")

ggplot(review_scores, aes(x = Year, y = score, group = review_type, color = review_type)) +
  geom_line() +
  ggtitle("Average Review Scores by Year")

```

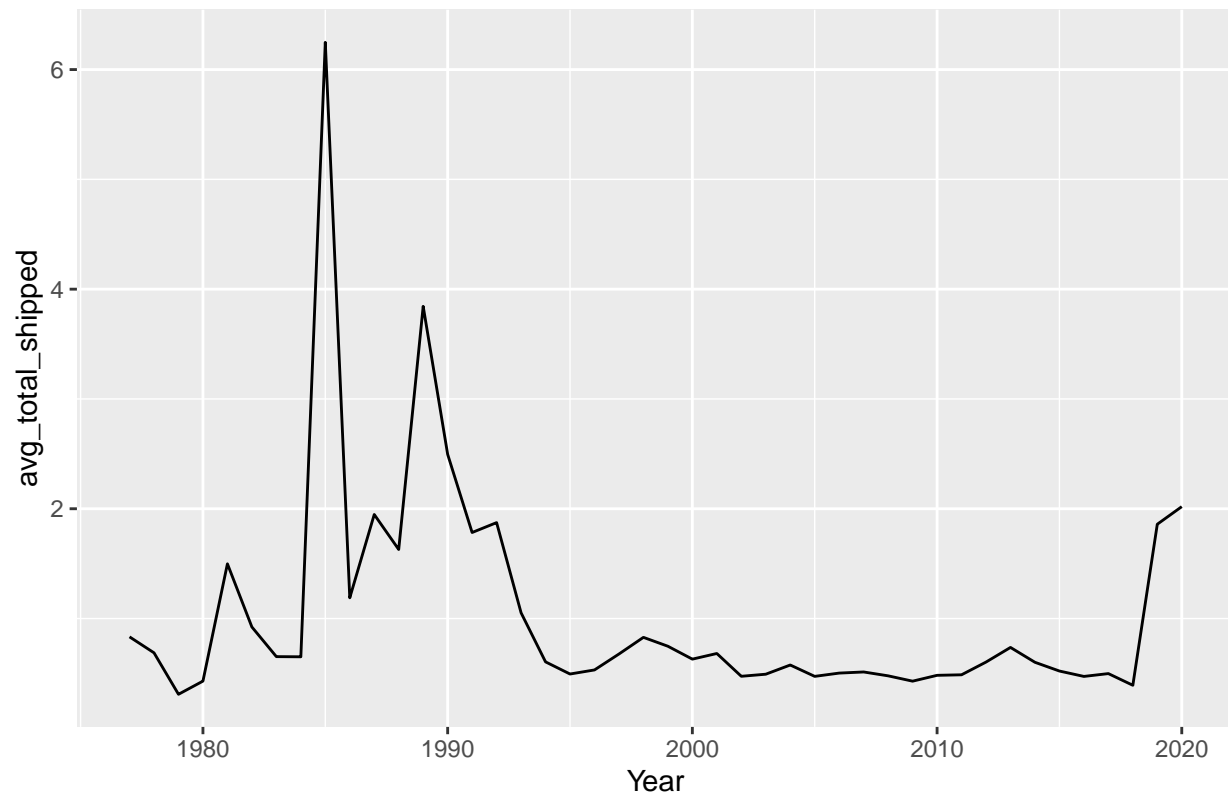


```

ggplot(vgs_year_s, aes(x = Year, y = avg_total_shipped)) +
  geom_line() +
  ggtitle("Average Shipments of Game Units by Year")

```

Average Shipments of Game Units by Year



```
# There are some major peaks in 1985 and 1989. Let's find out
# what cause these sudden surges in game shipment numbers.
vgs[which(vgs$Year == 1985), c(2, 8, 9)]
```

```
##           Name Total_Shipped Year
## 2      Super Mario Bros.      40.24 1985
## 13         Duck Hunt      28.31 1985
## 364       Excitebike       4.16 1985
## 386         Golf       4.01 1985
## 473        Kung Fu       3.50 1985
## 532       Baseball       3.20 1985
## 943        Tennis       2.17 1985
## 1161       Pinball       1.85 1985
## 1516      Ice Climber       1.50 1985
## 1774      Gyromite       1.32 1985
## 1856     Hogan's Alley       1.27 1985
## 2954     Clu Clu Land       0.82 1985
## 4502     Wrecking Crew       0.51 1985
## 4969     Ghostbusters       0.45 1985
## 5396        Dig Dug       0.41 1985
```

```
vgs[which(vgs$Year == 1989), c(2, 8, 9)]
```

```
##           Name Total_Shipped Year
## 11         Tetris      30.26 1989
```

```
## 28          Super Mario Land      18.14 1989
## 215          Tetris                5.58 1989
## 362 Teenage Mutant Ninja Turtles  4.17 1989
## 1015 Famista '89 - Kaimaku Han!!  2.05 1989
## 1052          Dragon Warrior      2.00 1989
## 1070          Tennis              1.99 1989
## 1100          Alleyway            1.94 1989
## 1326          Disney's DuckTales  1.67 1989
## 1378          Baseball            1.61 1989
## 1464          Tecmo Bowl          1.53 1989
## 1495          Mega Man 2          1.51 1989
## 1837          Yakuman             1.28 1989
## 2185 Famicom Jump: Eiyuu Retsuden 1.10 1989
## 2371          Bomberman           1.03 1989
## 4778          Double Dragon       0.47 1989
## 7889          Rampage              0.24 1989
## 8364          Ghostbusters II     0.21 1989
## 14400         Adventures of Lolo   0.06 1989
## 16864         SimCity              0.03 1989
```

*# As we can see, these 2 years have a low samples size and
were also headlined by major video game releases, namely
Super Mario Bros and Duck Hunt in 1985 followed by
Tetris and Super Mario Land in 1989, causing the
average total shipping numbers to be inflated.*

Observing the Discrepancy in Critic and User Review Scores

```
# Critic_Score - User_Score = Difference Score
dif <- vgs$Critic_Score - vgs$User_Score
na_index <- which(is.na(dif) == TRUE)
dif_df <- cbind("game" = vgs$Name, vgs[, c(6, 7)], "difference" = dif)
dif_df1 <- dif_df[-na_index, ]

# Games favored more positively by the critic than the user
head(dif_df1 %>% arrange(desc(difference)), 25)
```

```
##          game Critic_Score User_Score difference
## 1          FIFA 20          7.9         1.1         6.8
## 2    Call of Duty: Ghosts          8.8         2.1         6.7
## 3          FIFA 19          8.3         1.7         6.6
## 4        NBA 2K20          7.8         1.2         6.6
## 5        NBA 2K18          8.1         1.7         6.4
## 6          FIFA 19          8.3         2.0         6.3
## 7    Company of Heroes 2          8.4         2.1         6.3
## 8        NBA 2K18          7.7         1.5         6.2
## 9        Madden NFL 20          7.6         1.5         6.1
## 10 Star Wars Battlefront II          7.0         1.0         6.0
## 11        Madden NFL 19          8.0         2.3         5.7
## 12        NBA 2K19          8.2         2.6         5.6
## 13        NBA 2K19          8.3         2.8         5.5
```

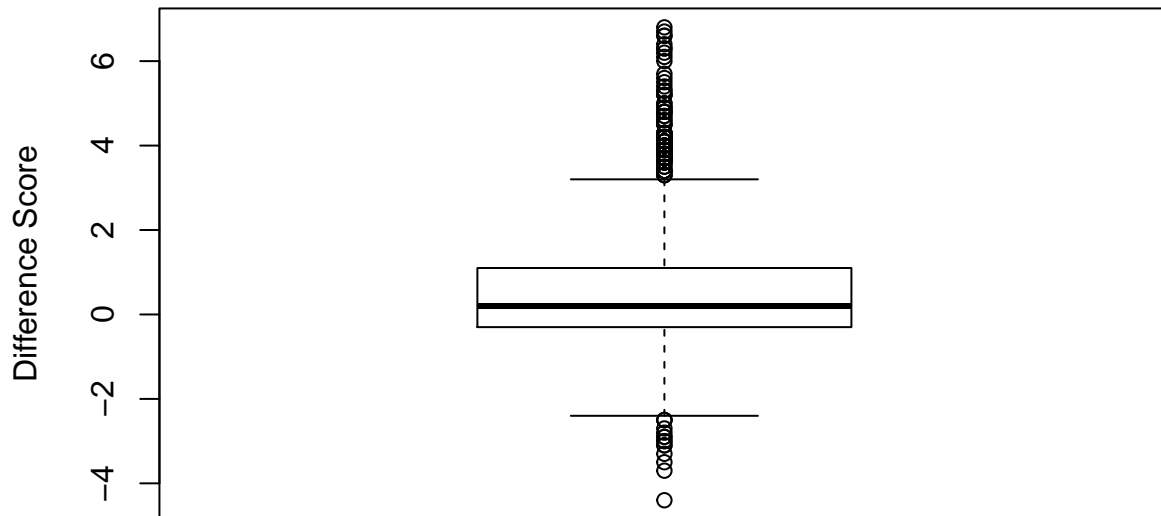
## 14	FIFA 18	8.0	2.6	5.4
## 15	Call of Duty: Modern Warfare 3	8.0	2.7	5.3
## 16	Battlefield V	8.1	2.8	5.3
## 17	Battlefield V	7.3	2.1	5.2
## 18	Battlefield V	7.8	2.6	5.2
## 19	Fortnite	8.1	2.9	5.2
## 20	Dragon Quest VII	8.0	3.0	5.0
## 21	FIFA 18	8.0	3.0	5.0
## 22	Diablo III	9.0	4.1	4.9
## 23	FIFA 18	8.3	3.4	4.9
## 24	Call of Duty: Black Ops II	9.3	4.4	4.9
## 25	World of Warcraft: Battle for Azeroth	7.9	3.1	4.8

```
# Games favored more positively by the user than the critic
tail(dif_df1 %>% arrange(desc(difference)), 25)
```

##	game	Critic_Score	User_Score	difference
## 2184	Valhalla Knights 3	4.8	6.9	-2.1
## 2185	Sonic Boom: Shattered Crystal	3.8	6.0	-2.2
## 2186	Borderlands 2	5.4	7.7	-2.3
## 2187	Fire Emblem Warriors	7.3	9.6	-2.3
## 2188	Crash Bandicoot N. Sane Trilogy	6.0	8.3	-2.3
## 2189	Toy Soldiers: War Chest	6.0	8.3	-2.3
## 2190	Samurai Warriors Chronicles	5.4	7.8	-2.4
## 2191	Lunar: Silver Star Story Complete	7.6	10.0	-2.4
## 2192	Has-Been Heroes	4.8	7.2	-2.4
## 2193	The Haunted: Hell's Reach	5.3	7.7	-2.4
## 2194	Just Dance 4	5.3	7.8	-2.5
## 2195	One Piece: Romance Dawn	4.2	6.7	-2.5
## 2196	Penny-Punching Princess	6.5	9.0	-2.5
## 2197	Crash of the Titans	6.5	9.2	-2.7
## 2198	Ragnarok Odyssey	5.0	7.8	-2.8
## 2199	Castlevania Judgment	4.6	7.5	-2.9
## 2200	Submarine Titans	6.1	9.0	-2.9
## 2201	Golden Axe: Beast Rider	4.0	7.0	-3.0
## 2202	Deadfall Adventures	4.0	7.0	-3.0
## 2203	Nickelodeon Kart Racers	2.0	5.1	-3.1
## 2204	Dead or Alive Xtreme 3: Fortune	4.3	7.4	-3.1
## 2205	FIFA 18	5.8	9.1	-3.3
## 2206	Past Cure	3.4	6.9	-3.5
## 2207	Sonic and the Black Knight	4.8	8.5	-3.7
## 2208	Valhalla Knights: Eldar Saga	3.5	7.9	-4.4

```
boxplot(dif_df1$difference, main = "Boxplot of Review Discrepancy",
        ylab = "Difference Score")
```

Boxplot of Review Discrepancy



```
summary(dif_df1$difference)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -4.4000 -0.3000  0.2000  0.4812  1.1000  6.8000
```

```
no_dif <- which(dif_df1$difference == 0)
# Rare occasions where critic and user reviews agree on a score.
tail(dif_df1[no_dif, ], 20)
```

##	game	Critic_Score	User_Score	difference
## 14704	Black Mirror	5.4	5.4	0
## 15110	Yomawari: The Long Night Collection	7.3	7.3	0
## 15242	Assault Suit Leynos	6.7	6.7	0
## 15447	Runaway: A Twist of Fate	7.9	7.9	0
## 15486	Darksiders: Warmastered Edition	8.0	8.0	0
## 15499	Valkyria Revolution	5.6	5.6	0
## 15712	Aragami	7.1	7.1	0
## 15767	Tour de France 2014	5.3	5.3	0
## 16115	Metal Gear Solid: Snake Eater 3D	7.9	7.9	0
## 16383	World of Final Fantasy Maxima	8.1	8.1	0
## 16502	Don't Knock Twice	5.6	5.6	0
## 16696	Death end re;Quest	7.4	7.4	0
## 16893	Magicka	7.6	7.6	0
## 16913	Root Letter	6.8	6.8	0
## 17046	Syberia	7.6	7.6	0

## 17547	Oil Rush	6.5	6.5	0
## 18311	Let It Die	7.2	7.2	0
## 18496	Battle Worlds: Kronos	7.1	7.1	0
## 18599	Titan Quest	6.5	6.5	0
## 18783	Transformers: Fall of Cybertron	7.8	7.8	0

*# Often these discrepancies in scores for the two types
are due to limitations of professionalism in critic reviews
versus the lack of in the average user reviews. Sometimes,
games are "review-bombed", a trend where ordinary internet
users down user reviews of a game due to a variety of reasons
such as disagreements in publisher/developer policies and ideals.
We can see that in the FIFA games where due to user complaints
of the games' annual releases with little changes, we see a huge
difference in the way a user and a critic reviews these games.
Often, they are just the work of disagreements in the quality of
the game; maybe a critic just wasn't able to see the positives of
a certain game and reviewed it unfavorably while users loved the game.*

Comparing Publishers by Total Number of Game Units Shipped

```
vgs_pub <- vgs %>%
  select(Publisher, Critic_Score, Total_Shipped) %>%
  group_by(Publisher) %>%
  summarise(
    overall_total = sum(Total_Shipped, na.rm = T),
    avg_c_reviews = mean(Critic_Score, na.rm = T)
  )
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

A list of the most successful publishers saleswise.

```
head(vgs_pub %>% arrange(desc(overall_total)), 10)
```

```
## # A tibble: 10 x 3
##   Publisher          overall_total avg_c_reviews
##   <fct>              <dbl>         <dbl>
## 1 Nintendo          2095.         7.72
## 2 Activision         760.         6.88
## 3 Electronic Arts    687.         7.32
## 4 Sony Computer Entertainment 568.         7.51
## 5 EA Sports          501.         7.89
## 6 Ubisoft            499.         6.91
## 7 THQ                342.         6.76
## 8 Sega               305.         7.17
## 9 Rockstar Games     274.         8.43
## 10 Capcom            271.         7.44
```



```

# A list of publishers with most critically acclaimed games.
# Keep in mind most of these publishers are indie w/ only one or slightly more
# games under their belt, which is why most of these publishers
# are not familiar,]. regardless, their games became major hits.
# (Also, most of these publishers sell digitally, explaining their
# low physical shipment/sales total)
head(vgs_pub %>% arrange(desc(avg_c_reviews)), 20)

```

```

## # A tibble: 20 x 3
##   Publisher      overall_total avg_c_reviews
##   <fct>          <dbl>         <dbl>
## 1 Trapdoor        1.28           9.5
## 2 Facepalm Games  0.48           9.3
## 3 Chucklefish     4.91           9.2
## 4 Psyonix Studios 0.12           9.2
## 5 Toby Fox        3.6            9.2
## 6 Valve          74.8           9.2
## 7 Runic Games      3            9.1
## 8 Valve Software  3.28           9.1
## 9 4-Aug           0.02           9
## 10 Matt Makes Games Inc. 0.21           9
## 11 Messhof        0.79           9
## 12 Mossmouth, LLC 0.91           9
## 13 Rare           4.45           9
## 14 Villa Gorilla   0.04           9
## 15 Mega Crit Games 1.12           8.9
## 16 Motion Twin     0.83           8.9
## 17 Obsidian Entertainment 0.02           8.9
## 18 Orbital Media, Inc. 0.06           8.9
## 19 Sierra Studios  4.2            8.9
## 20 ESP            0.54           8.85

```

```

# Just for fun. let's see what the critic review score average for
# Nintendo is
vgs_pub[which(vgs_pub$Publisher == "Nintendo"), ]

```

```

## # A tibble: 1 x 3
##   Publisher overall_total avg_c_reviews
##   <fct>          <dbl>         <dbl>
## 1 Nintendo    2095.           7.72

```