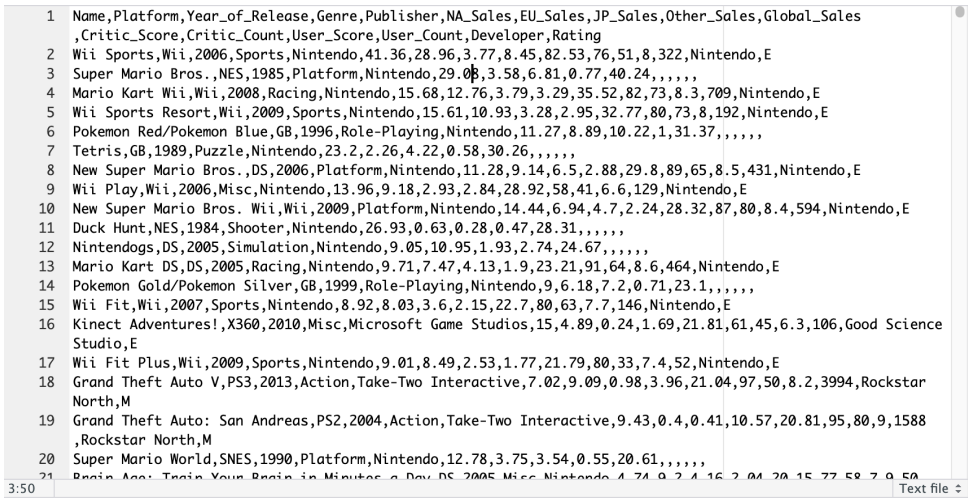
ENG WEI KHUAN
December 14 2024
2024 FALL
STAT 419-10
Prof. Renae Shrum

Video game sales with rating

Before the first step, I download the csv from a website that I plan to use for my project (Kaggle).

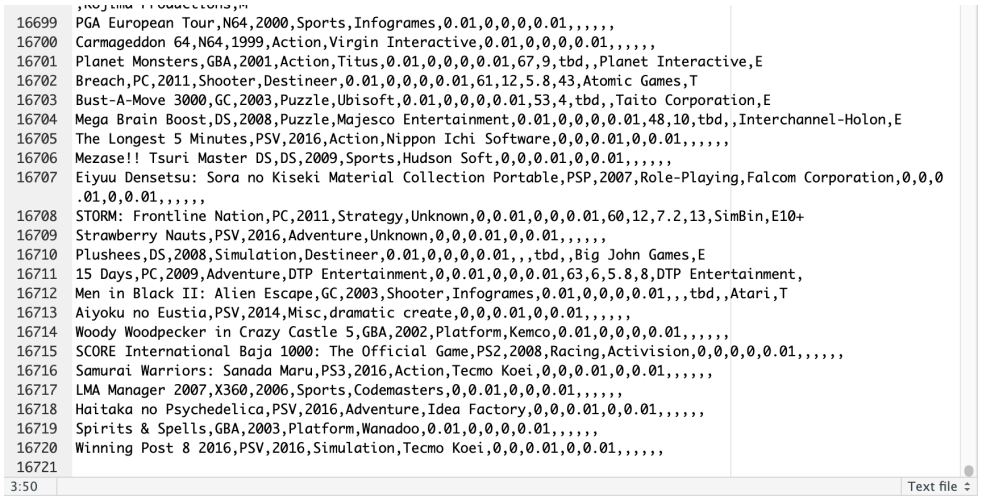
The folder was named “Video_Games_Sales_as_at_22_Dec_2016.csv”, and I named it “VG.csv” because I think it’s too long.

Below is how the data looks like in “Video_Games_Sales_as_at_22_Dec_2016.csv” or “VG.csv”



```
1  Name,Platform,Year_of_Release,Genre,Publisher,NA_Sales,EU_Sales,JP_Sales,Other_Sales,Global_Sales
   ,Critic_Score,Critic_Count,User_Score,User_Count,Developer,Rating
2  Wii Sports,Wii,2006,Sports,Nintendo,41.36,28.96,3.77,8.45,82.53,76,51,8,322,Nintendo,E
3  Super Mario Bros.,NES,1985,Platform,Nintendo,29.08,3.58,6.81,0.77,40.24,,,,,,
4  Mario Kart Wii,Wii,2008,Racing,Nintendo,15.68,12.76,3.79,3.29,35.52,82,73,8.3,709,Nintendo,E
5  Wii Sports Resort,Wii,2009,Sports,Nintendo,15.61,10.93,3.28,2.95,32.77,80,73,8,192,Nintendo,E
6  Pokemon Red/Pokemon Blue,GB,1996,Role-Playing,Nintendo,11.27,8.89,10.22,1,31.37,,,,,,
7  Tetris,GB,1989,Puzzle,Nintendo,23.2,2.26,4.22,0.58,30.26,,,,,,
8  New Super Mario Bros.,DS,2006,Platform,Nintendo,11.28,9.14,6.5,2.88,29.8,89,65,8.5,431,Nintendo,E
9  Wii Play,Wii,2006,Misc,Nintendo,13.96,9.18,2.93,2.84,28.92,58,41,6.6,129,Nintendo,E
10 New Super Mario Bros. Wii,Wii,2009,Platform,Nintendo,14.44,6.94,4.7,2.24,28.32,87,80,8.4,594,Nintendo,E
11 Duck Hunt,NES,1984,Shooter,Nintendo,26.93,0.63,0.28,0.47,28.31,,,,,,
12 Nintendogs,DS,2005,Simulation,Nintendo,9.05,10.95,1.93,2.74,24.67,,,,,,
13 Mario Kart DS,DS,2005,Racing,Nintendo,9.71,7.47,4.13,1.9,23.21,91,64,8.6,464,Nintendo,E
14 Pokemon Gold/Pokemon Silver,GB,1999,Role-Playing,Nintendo,9,6.18,7.2,0.71,23.1,,,,,,
15 Wii Fit,Wii,2007,Sports,Nintendo,8.92,8.03,3.6,2.15,22.7,80,63,7.7,146,Nintendo,E
16 Kinect Adventures!,X360,2010,Misc,Microsoft Game Studios,15,4.89,0.24,1.69,21.81,61,45,6.3,106,Good Science
   Studio,E
17 Wii Fit Plus,Wii,2009,Sports,Nintendo,9.01,8.49,2.53,1.77,21.79,80,33,7.4,52,Nintendo,E
18 Grand Theft Auto V,PS3,2013,Action,Take-Two Interactive,7.02,9.09,0.98,3.96,21.04,97,50,8.2,3994,Rockstar
   North,M
19 Grand Theft Auto: San Andreas,PS2,2004,Action,Take-Two Interactive,9.43,0.4,0.41,10.57,20.81,95,80,9,1588
   ,Rockstar North,M
20 Super Mario World,SNES,1990,Platform,Nintendo,12.78,3.75,3.54,0.55,20.61,,,,,,
21 Brain Age: Train Your Brain in Minutes a Day,DS,2005,Misc,Nintendo,4.74,9.2,4.16,2.04,20.15,77,58,7.9,50
```

3:50 Text file



```
16699 PGA European Tour,N64,2000,Sports,Infogrames,0.01,0,0,0,0.01,,,,,,
16700 Carmageddon 64,N64,1999,Action,Virgin Interactive,0.01,0,0,0,0.01,,,,,,
16701 Planet Monsters,GBA,2001,Action,Titus,0.01,0,0,0,0.01,67,9,tbd,,Planet Interactive,E
16702 Breach,PC,2011,Shooter,Destineer,0.01,0,0,0,0.01,61,12,5.8,43,Atomic Games,T
16703 Bust-A-Move 3000,GC,2003,Puzzle,Ubisoft,0.01,0,0,0,0.01,53,4,tbd,,Taito Corporation,E
16704 Mega Brain Boost,DS,2008,Puzzle,Majesco Entertainment,0.01,0,0,0,0.01,48,10,tbd,,Interchannel-Holon,E
16705 The Longest 5 Minutes,PSV,2016,Action,Nippon Ichi Software,0,0,0.01,0,0.01,,,,,,
16706 Mezase!! Tsuri Master DS,DS,2009,Sports,Hudson Soft,0,0,0.01,0,0.01,,,,,,
16707 Eiyuu Densetsu: Sora no Kiseki Material Collection Portable,PSP,2007,Role-Playing,Falcom Corporation,0,0,0
      .01,0,0.01,,,,,,
16708 STORM: Frontline Nation,PC,2011,Strategy,Unknown,0,0.01,0,0,0.01,60,12,7.2,13,SimBin,E10+
16709 Strawberry Nauts,PSV,2016,Adventure,Unknown,0,0,0.01,0,0.01,,,,,,
16710 Plushees,DS,2008,Simulation,Destineer,0.01,0,0,0,0.01,,,tbd,,Big John Games,E
16711 15 Days,PC,2009,Adventure,DTP Entertainment,0,0.01,0,0,0.01,63,6,5.8,8,DTP Entertainment,
16712 Men in Black II: Alien Escape,GC,2003,Shooter,Infogrames,0.01,0,0,0,0.01,,,tbd,,Atari,T
16713 Aiyoku no Eustia,PSV,2014,Misc,dramatic create,0,0,0.01,0,0.01,,,,,,
16714 Woody Woodpecker in Crazy Castle 5,GBA,2002,Platform,Kemco,0.01,0,0,0,0.01,,,,,,
16715 SCORE International Baja 1000: The Official Game,PS2,2008,Racing,Activision,0,0,0,0,0.01,,,,,,
16716 Samurai Warriors: Sanada Maru,PS3,2016,Action,Tecmo Koei,0,0,0.01,0,0.01,,,,,,
16717 LMA Manager 2007,X360,2006,Sports,Codemasters,0,0.01,0,0,0.01,,,,,,
16718 Haitaka no Psychedelica,PSV,2016,Adventure,Idea Factory,0,0,0.01,0,0.01,,,,,,
16719 Spirits & Spells,GBA,2003,Platform,Wanadoo,0.01,0,0,0,0.01,,,,,,
16720 Winning Post 8 2016,PSV,2016,Simulation,Tecmo Koei,0,0,0.01,0,0.01,,,,,,
16721
```

3:50 Text file

There is some missing info and there’s 16,720 lines of info.

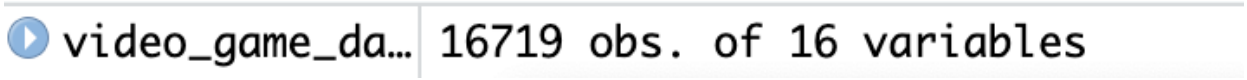
On the next slide, I will read those data into my Rstudio and make it look more clean.

1) video_game_data <- read.csv("VG.csv", header = TRUE)

	Name	Platform	Year_of_Release	Genre	Publisher	NA_Sales	EU_Sales	JP_Sales	Other_Sales	Global_Sales	Critic_Scc
2009	Wii Sports	Wii	2006	Sports	Nintendo	41.36	28.96	3.77	8.45	82.53	
2	Super Mario Bros.	NES	1985	Platform	Nintendo	29.08	3.58	6.81	0.77	40.24	
3	Mario Kart Wii	Wii	2008	Racing	Nintendo	15.68	12.76	3.79	3.29	35.52	
4	Wii Sports Resort	Wii	2009	Sports	Nintendo	15.61	10.93	3.28	2.95	32.77	
5	Pokemon Red/Pokemon Blue	GB	1996	Role-Playing	Nintendo	11.27	8.89	10.22	1.00	31.37	
6	Tetris	GB	1989	Puzzle	Nintendo	23.20	2.26	4.22	0.58	30.26	
7	New Super Mario Bros.	DS	2006	Platform	Nintendo	11.28	9.14	6.50	2.88	29.80	
8	Wii Play	Wii	2006	Misc	Nintendo	13.96	9.18	2.93	2.84	28.92	
9	New Super Mario Bros. Wii	Wii	2009	Platform	Nintendo	14.44	6.94	4.70	2.24	28.32	
10	Duck Hunt	NES	1984	Shooter	Nintendo	26.93	0.63	0.28	0.47	28.31	
11	Nintendogs	DS	2005	Simulation	Nintendo	9.05	10.95	1.93	2.74	24.67	
12	Mario Kart DS	DS	2005	Racing	Nintendo	9.71	7.47	4.13	1.90	23.21	
13	Pokemon Gold/Pokemon Silver	GB	1999	Role-Playing	Nintendo	9.00	6.18	7.20	0.71	23.10	
14	Wii Fit	Wii	2007	Sports	Nintendo	8.92	8.03	3.60	2.15	22.70	
15	Kinect Adventures!	X360	2010	Misc	Microsoft Game Studios	15.00	4.89	0.24	1.69	21.81	
16	Wii Fit Plus	Wii	2009	Sports	Nintendo	9.01	8.49	2.53	1.77	21.79	
17	Grand Theft Auto V	PS3	2013	Action	Take-Two Interactive	7.02	9.09	0.98	3.96	21.04	
18	Grand Theft Auto: San Andreas	PS2	2004	Action	Take-Two Interactive	9.43	0.40	0.41	10.57	20.81	
19	Super Mario World	SNES	1990	Platform	Nintendo	12.78	3.75	3.54	0.55	20.61	
20	Brain Age: Train Your Brain in Minutes a Day	DS	2005	Misc	Nintendo	4.74	9.20	4.16	2.04	20.15	
21	Pokemon Diamond/Pokemon Pearl	DS	2006	Role-Playing	Nintendo	6.38	4.46	6.04	1.36	18.25	
22	Super Mario Land	GB	1989	Platform	Nintendo	10.83	2.71	4.18	0.42	18.14	
23	Super Mario Bros. 3	NES	1988	Platform	Nintendo	9.54	3.44	3.84	0.46	17.28	
24	Grand Theft Auto V	X360	2013	Action	Take-Two Interactive	9.66	5.14	0.06	1.41	16.27	
25	Grand Theft Auto: Vice City	PS2	2002	Action	Take-Two Interactive	8.41	5.49	0.47	1.78	16.15	
26	Pokemon Ruby/Pokemon Sapphire	GBA	2002	Role-Playing	Nintendo	6.06	3.90	5.38	0.50	15.85	
27	Brain Age 2: More Training in Minutes a Day	DS	2005	Puzzle	Nintendo	3.43	5.35	5.32	1.18	15.29	
28	Pokemon Black/Pokemon White	DS	2010	Role-Playing	Nintendo	5.51	3.17	5.65	0.80	15.14	
29	Gran Turismo 3: A-Spec	PS2	2001	Racing	Sony Computer Entertainment	6.85	5.09	1.87	1.16	14.98	
30	Call of Duty: Modern Warfare 3	X360	2011	Shooter	Activision	9.04	4.24	0.13	1.32	14.73	

NA_Sales	EU_Sales	JP_Sales	Other_Sales	Global_Sales	Critic_Score	Critic_Count	User_Score	User_Count	Developer	Rating
41.36	28.96	3.77	8.45	82.53	76	51	8	322	Nintendo	E
29.08	3.58	6.81	0.77	40.24	NA	NA		NA		
15.68	12.76	3.79	3.29	35.52	82	73	8.3	709	Nintendo	E
15.61	10.93	3.28	2.95	32.77	80	73	8	192	Nintendo	E
11.27	8.89	10.22	1.00	31.37	NA	NA		NA		
23.20	2.26	4.22	0.58	30.26	NA	NA		NA		
11.28	9.14	6.50	2.88	29.80	89	65	8.5	431	Nintendo	E
13.96	9.18	2.93	2.84	28.92	58	41	6.6	129	Nintendo	E
14.44	6.94	4.70	2.24	28.32	87	80	8.4	594	Nintendo	E
26.93	0.63	0.28	0.47	28.31	NA	NA		NA		
9.05	10.95	1.93	2.74	24.67	NA	NA		NA		
9.71	7.47	4.13	1.90	23.21	91	64	8.6	464	Nintendo	E
9.00	6.18	7.20	0.71	23.10	NA	NA		NA		
8.92	8.03	3.60	2.15	22.70	80	63	7.7	146	Nintendo	E
15.00	4.89	0.24	1.69	21.81	61	45	6.3	106	Good Science Studio	E
9.01	8.49	2.53	1.77	21.79	80	33	7.4	52	Nintendo	E
7.02	9.09	0.98	3.96	21.04	97	50	8.2	3994	Rockstar North	M
9.43	0.40	0.41	10.57	20.81	95	80	9	1588	Rockstar North	M
12.78	3.75	3.54	0.55	20.61	NA	NA		NA		
4.74	9.20	4.16	2.04	20.15	77	58	7.9	50	Nintendo	E
6.38	4.46	6.04	1.36	18.25	NA	NA		NA		
10.83	2.71	4.18	0.42	18.14	NA	NA		NA		
9.54	3.44	3.84	0.46	17.28	NA	NA		NA		

With this code, everything look much more clean



While reading the data, we can see they're 16,719 observation

2) head(video_game_data)
   str(video_game_data)

```
> head(video_game_data)
                      Name Platform Year_of_Release         Genre Publisher NA_Sales EU_Sales JP_Sales Other_Sales Global_Sales
1               Wii Sports      Wii            2006        Sports  Nintendo    41.36    28.96     3.77        8.45        82.53
2        Super Mario Bros.      NES            1985      Platform  Nintendo    29.08     3.58     6.81        0.77        40.24
3           Mario Kart Wii      Wii            2008        Racing  Nintendo    15.68    12.76     3.79        3.29        35.52
4        Wii Sports Resort      Wii            2009        Sports  Nintendo    15.61    10.93     3.28        2.95        32.77
5 Pokemon Red/Pokemon Blue       GB            1996 Role-Playing  Nintendo    11.27     8.89    10.22        1.00        31.37
6                  Tetris       GB            1989        Puzzle  Nintendo    23.20     2.26     4.22        0.58        30.26
  Critic_Score Critic_Count User_Score User_Count Developer Rating
1           76           51          8        322  Nintendo      E
2           NA           NA                    NA
3           82           73        8.3        709  Nintendo      E
4           80           73          8        192  Nintendo      E
5           NA           NA                    NA
6           NA           NA                    NA
```

```
> str(video_game_data)# Get the structure of the dataset
'data.frame':   16719 obs. of  16 variables:
 $ Name           : chr  "Wii Sports" "Super Mario Bros." "Mario Kart Wii" "Wii Sports Resort" ...
 $ Platform       : chr  "Wii" "NES" "Wii" "Wii" ...
 $ Year_of_Release: chr  "2006" "1985" "2008" "2009" ...
 $ Genre          : chr  "Sports" "Platform" "Racing" "Sports" ...
 $ Publisher      : chr  "Nintendo" "Nintendo" "Nintendo" "Nintendo" ...
 $ NA_Sales       : num  41.4 29.1 15.7 15.6 11.3 ...
 $ EU_Sales       : num  28.96 3.58 12.76 10.93 8.89 ...
 $ JP_Sales       : num  3.77 6.81 3.79 3.28 10.22 ...
 $ Other_Sales    : num  8.45 0.77 3.29 2.95 1 0.58 2.88 2.84 2.24 0.47 ...
 $ Global_Sales   : num  82.5 40.2 35.5 32.8 31.4 ...
 $ Critic_Score   : int  76 NA 82 80 NA NA 89 58 87 NA ...
 $ Critic_Count   : int  51 NA 73 73 NA NA 65 41 80 NA ...
 $ User_Score     : chr  "8" "" "8.3" "8" ...
 $ User_Count     : int  322 NA 709 192 NA NA 431 129 594 NA ...
 $ Developer      : chr  "Nintendo" "" "Nintendo" "Nintendo" ...
 $ Rating         : chr  "E" "" "E" "E" ...
>
```

Above two graphs show their structure and header.

3) summary(video_game_data)

```
> summary(video_game_data)
     Name             Platform         Year_of_Release       Genre             Publisher           NA_Sales
 Length:16719       Length:16719       Length:16719       Length:16719       Length:16719       Min.   : 0.0000
 Class :character   Class :character   Class :character   Class :character   Class :character   1st Qu.: 0.0000
 Mode  :character   Mode  :character   Mode  :character   Mode  :character   Mode  :character   Median : 0.0800
                                                                                                Mean   : 0.2633
                                                                                                3rd Qu.: 0.2400
                                                                                                Max.   :41.3600

    EU_Sales         JP_Sales        Other_Sales        Global_Sales      Critic_Score    Critic_Count     User_Score
 Min.   : 0.000   Min.   : 0.0000   Min.   : 0.00000   Min.   : 0.0100   Min.   :13.00   Min.   :  3.00   Length:16719
 1st Qu.: 0.000   1st Qu.: 0.0000   1st Qu.: 0.00000   1st Qu.: 0.0600   1st Qu.:60.00   1st Qu.: 12.00   Class :character
 Median : 0.020   Median : 0.0000   Median : 0.01000   Median : 0.1700   Median :71.00   Median : 21.00   Mode  :character
 Mean   : 0.145   Mean   : 0.0776   Mean   : 0.04733   Mean   : 0.5335   Mean   :68.97   Mean   : 26.36
 3rd Qu.: 0.110   3rd Qu.: 0.0400   3rd Qu.: 0.03000   3rd Qu.: 0.4700   3rd Qu.:79.00   3rd Qu.: 36.00
 Max.   :28.960   Max.   :10.2200   Max.   :10.57000   Max.   :82.5300   Max.   :98.00   Max.   :113.00
                                                                        NA's   :8582    NA's   :8582

  User_Count       Developer           Rating
 Min.   :    4.0   Length:16719       Length:16719
 1st Qu.:   10.0   Class :character   Class :character
 Median :   24.0   Mode  :character   Mode  :character
 Mean   :  162.2
 3rd Qu.:   81.0
 Max.   :10665.0
 NA's   :9129
```

Summary of my dataset. From here we can see the IQR of all headers.

But we can see some missing values (n/a), so in my next few slides I would like to deal with missing values(n/a).

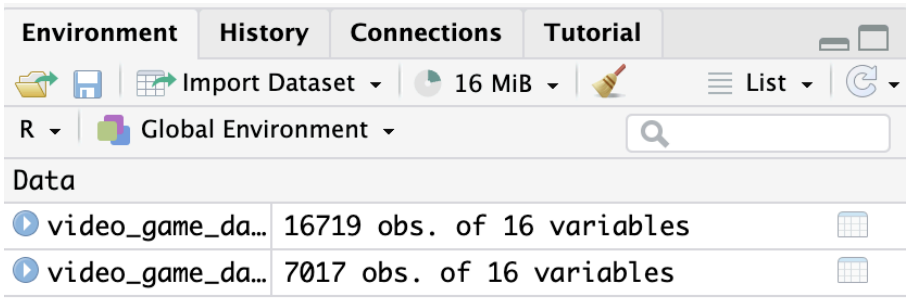
4) colSums(is.na(video_game_data))

```
> colSums(is.na(video_game_data))# Check for missing values
          Name       Platform Year_of_Release          Genre      Publisher       NA_Sales       EU_Sales       JP_Sales
             0              0              0              0              0              0              0              0
   Other_Sales   Global_Sales   Critic_Score   Critic_Count     User_Score     User_Count      Developer         Rating
             0              0           8582           8582              0           9129              0              0
>
```

With this code I can see missing value in specific column

5) video_game_data_clean <- video_game_data[!is.na(video_game_data$Critic_Score) & !is.na(video_game_data$User_Count),]

This code is to remove rows with missing values in key columns.



The first one(16,719 obs) is the data before I run the code (video_game_data), and the second data (7017 observation) is the one after I run the code(video_game_data_clean). And we should see missing values are all removed. As below:

NA_Sales	EU_Sales	JP_Sales	Other_Sales	Global_Sales	Critic_Score	Critic_Count	User_Score	User_Count	Developer	Rating
41.36	28.96	3.77	8.45	82.53	76	51	8	322	Nintendo	E
15.68	12.76	3.79	3.29	35.52	82	73	8.3	709	Nintendo	E
15.61	10.93	3.28	2.95	32.77	80	73	8	192	Nintendo	E
11.28	9.14	6.50	2.88	29.80	89	65	8.5	431	Nintendo	E
13.96	9.18	2.93	2.84	28.92	58	41	6.6	129	Nintendo	E
14.44	6.94	4.70	2.24	28.32	87	80	8.4	594	Nintendo	E
9.71	7.47	4.13	1.90	23.21	91	64	8.6	464	Nintendo	E
8.92	8.03	3.60	2.15	22.70	80	63	7.7	146	Nintendo	E
15.00	4.89	0.24	1.69	21.81	61	45	6.3	106	Good Science Studio	E
9.01	8.49	2.53	1.77	21.79	80	33	7.4	52	Nintendo	E
7.02	9.09	0.98	3.96	21.04	97	50	8.2	3994	Rockstar North	M
9.43	0.40	0.41	10.57	20.81	95	80	9	1588	Rockstar North	M
4.74	9.20	4.16	2.04	20.15	77	58	7.9	50	Nintendo	E
9.66	5.14	0.06	1.41	16.27	97	58	8.1	3711	Rockstar North	M
8.41	5.49	0.47	1.78	16.15	95	62	8.7	730	Rockstar North	M
3.43	5.35	5.32	1.18	15.29	77	37	7.1	19	Nintendo	E
6.85	5.09	1.87	1.16	14.98	95	54	8.4	314	Polyphony Digital	E
9.04	4.24	0.13	1.32	14.73	88	81	3.4	8713	Infinity Ward, Sledgehammer Games	M
9.70	3.68	0.11	1.13	14.61	87	89	6.3	1454	Treyarch	M
4.99	5.73	0.65	2.42	13.79	83	21	5.3	922	Treyarch	M
8.25	4.24	0.07	1.12	13.67	83	73	4.8	2256	Treyarch	M
8.52	3.59	0.08	1.28	13.47	94	100	6.3	2698	Infinity Ward	M
5.54	5.73	0.49	1.57	13.32	88	39	3.2	5234	Infinity Ward, Sledgehammer Games	M
6.99	4.51	0.30	1.30	13.10	97	56	8.5	664	DMA Design	M
6.62	2.55	2.66	1.01	12.84	93	81	8.9	1662	Game Arts	T
5.03	4.02	2.69	0.91	12.66	85	73	8.2	632	Retro Studios, Entertainment Analysis & Development…	E

6) aov_sales <- aov(Global_Sales ~ Genre, data = video_game_data)

aov_sales	Large aov (13 elements, 5.8 MB)
aov_sales	list [13] (S3: aov, lm)	List of length 13
coefficients	double [13]	1.210 −0.692 −1.028 −0.683 −0.751 −0.277 ...
residuals	double [16719]	82.0 39.3 34.9 32.2 30.7 29.8 ...
effects	double [16719]	−68.99 −1.02 −13.49 −1.48 −5.49 10.35 ...
rank	integer [1]	13
fitted.values	double [16719]	0.567 0.933 0.584 0.567 0.623 0.419 ...
assign	integer [13]	0 1 1 1 1 1 ...
qr	list [5] (S3: qr)	List of length 5
df.residual	integer [1]	16706
contrasts	list [1]	List of length 1
xlevels	list [1]	List of length 1
call	language	aov(formula = Global_Sales ~ Genre, data = video_game_data)
terms	formula	Global_Sales ~ Genre
model	list [16719 x 2] (S3: data.fram	A data.frame with 16719 rows and 2 columns

This code performs an **Analysis of Variance (ANOVA)** to determine if there are significant differences in `Global_Sales` across different levels of `Genre` in the dataset `video_game_data`.

In this code: “aov_sales <- aov(Global_Sales ~ Genre, data = video_game_data)”, we are comparing the variation in Global_Sales across different Genre categories. We can replace Global_Sales with any numeric column (the dependent variable) we want to analyze, and Genre with any categorical column (factor) we want to use for comparison.
After performing the ANOVA, we use the summary(aov_sales) to examine the statistical results, which helps us apply statistical knowledge and make conclusions about the relationships between the variables.

7) Then, we perform **summary(aov_sales)**

```
              Df Sum Sq Mean Sq F value Pr(>F)
Genre         12    489   40.78   17.22 <2e-16 ***
Residuals   16706  39569    2.37
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> 
```

From **summary(aov_sales)** can see:

degree freedom: 13-1=12

F-value: 17.22,suggests significant differences between genres.

P-value: <2e-16, much smaller than 0.05 so it is statistically significant. It means there are significant differences in Global_Sales across genres.

—----------

ANOVA only checks for differences across the entire set of groups.
So I decided to make pairwise comparisons (**Tukey**, next slide).

8) tukey_test <- TukeyHSD(aov_sales)

tukey_test	list [1] (S3: TukeyHSD, multic	List of length 1
Genre	double [78 x 4]	-0.692 -1.028 -0.683 -0.751 -0.277 -0.791 -4.299 -4.636 -4.293 -4.359 -3.887 -4. ...

9) print(tukey_test)

```
> print(tukey_test)
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = Global_Sales ~ Genre, data = video_game_data)

$Genre
                              diff         lwr          upr     p adj
Action-                 -0.692115727 -4.29876725  2.914535797 0.9999897
Adventure-              -1.027582501 -4.63593036  2.580765353 0.9993112
Fighting-               -0.682932862 -4.29275900  2.926893274 0.9999912
Misc-                   -0.751040000 -4.35868152  2.856601521 0.9999746
Platform-               -0.277477477 -3.88711731  3.332162356 1.0000000
Puzzle-                 -0.791000000 -4.40279295  2.820792948 0.9999560
Racing-                 -0.626413130 -4.23488053  2.982054267 0.9999966
Role-Playing-           -0.587066666 -4.19505136  3.020918031 0.9999984
Shooter-                -0.414126984 -4.02243304  3.194179071 1.0000000
Simulation-             -0.763295194 -4.37299999  2.846409605 0.9999699
Sports-                 -0.642708688 -4.24982573  2.964408359 0.9999955
Strategy-               -0.954509516 -4.56536647  2.656347438 0.9996775
Adventure-Action        -0.335466775 -0.50180845 -0.169125096 0.0000000
Fighting-Action          0.009182865 -0.18662328  0.204989009 1.0000000
Misc-Action             -0.058924273 -0.20916627  0.091317726 0.9862590
Platform-Action          0.414638250  0.22229731  0.606979193 0.0000000
Puzzle-Action           -0.098884273 -0.32810820  0.130339654 0.9699915
Racing-Action            0.065702596 -0.10321238  0.234617574 0.9871947
Role-Playing-Action      0.105049060 -0.05321935  0.263317474 0.5918648
Shooter-Action           0.277988743  0.11255627  0.443421212 0.0000021
Simulation-Action       -0.071179468 -0.26473579  0.122376858 0.9921525
Sports-Action            0.049407039 -0.08766486  0.186478937 0.9934408
Strategy-Action         -0.262393790 -0.47636402 -0.048423563 0.0033269
Fighting-Adventure       0.344649640  0.11975173  0.569547554 0.0000294
Misc-Adventure           0.276542502  0.08996361  0.463121397 0.0000691
Platform-Adventure       0.750105024  0.52821752  0.971992529 0.0000000
Puzzle-Adventure         0.236582502 -0.01794183  0.491106832 0.0992912
Racing-Adventure         0.401169371  0.19925034  0.603088404 0.0000000
Role-Playing-Adventure   0.440515835  0.24741504  0.633616629 0.0000000
```

diff= -0.692
Lwr and upr = -4, lower bound and upper bound of the confidence interval suggests that the true difference in sales could be as large as **-0.40 million**
P adj= 0.9999897, close to 1 so it's non-significant p-value, as there is no actual comparison

```
Role-Playing-Action      0.105049060 -0.05321935  0.263317474 0.5918648
```

With this row info, Role playing - Action, diff=0.105, upr=-0.05322,lwr=0.26332, padj=0.592.

I can conclude that :

Role-Playing games have a mean global sales difference of 0.105 million units (105,000 units) higher than Action games.

The confidence interval spans from -0.0532 million to 0.2633 million (263,300 units).

The p-value is much greater than 0.05, meaning the difference is not statistically significant.

Then, I can make such a conclusion for another row from Print(tukey) code.

```
video_game_data <- read.csv("VG.csv", header = TRUE)
head(video_game_data)
str(video_game_data)# Get the structure of the dataset
summary(video_game_data)
colSums(is.na(video_game_data))# Check for missing values
colSums(is.na(video_game_data))# Count missing values per column
video_game_data_clean <- video_game_data[!is.na(video_game_data$Critic_Score) &
!is.na(video_game_data$User_Count), ]#Remove rows with missing values in key columns
aov_sales <- aov(Global_Sales ~ Genre, data = video_game_data)
summary(aov_sales)
tukey_test <- TukeyHSD(aov_sales)
print(tukey_test)
summary(tukey_test)
plot(TukeyHSD(aov_sales))
```

Work Cites

Kirubi R. “Video game sales with rating.” Kaggle, 8 years ago
https://www.kaggle.com/datasets/rush4ratio/video-game-sales-with-ratings

THANK YOU!