

# Assignment 2

Irfan Kanat

March 8, 2018

This is an assignment to evaluate your learning in module two. For each learning activity there is a section in this document. Answer each question below the question and create an R Markdown document for submission.

For the rest of the exercises we will use *games.csv* in the data folder. Import the file into a data.frame called *games* to begin the exercise.

```
# Import necessary packages
library(ggplot2)
# Import games.csv from data folder
games <- read.csv("data/games.csv")
```

## 1 - Describe

### Visual Inspection

Show the top 6 rows of *games*.

```
head(games)
```

```
##   appID dateMonth   status   avgMin listPrice  reviews multi
## 1   280   201601 SuperStar 75.52385  9.986786 83.00000     0
## 2  1250   201601 SuperStar 2103.36076 19.986786 96.00000     1
## 3  2300   201601   Niche   33.73054  4.986786 90.00000     1
## 4  3340   201601   Niche  146.63755  4.986786 10.00000     0
## 5  3960   201601   Niche   98.04734  4.990000 36.78571     1
## 6  4770   201601 Popular  125.17656  4.986786 73.00000     1
```

### Descriptive Statistics

Use *summary* function to provide basic descriptive statistics for *games*.

```
summary(games)
```

```
##      appID      dateMonth      status      avgMin
## Min.   : 280   Min.   :201601   Niche    :323   Min.   : 0.00
## 1st Qu.:205796 1st Qu.:201601   Popular : 26   1st Qu.: 0.00
## Median :263040 Median :201601   SuperStar: 9   Median : 80.86
## Mean   :231540 Mean   :201601               Mean   : 274.98
## 3rd Qu.:315542 3rd Qu.:201601               3rd Qu.: 223.34
## Max.   :423880 Max.   :201601               Max.   :9365.34
##
##      listPrice      reviews      multi
## Min.   : 0.000   Min.   :10.00   Min.   :0.0000
## 1st Qu.: 4.987   1st Qu.:57.46   1st Qu.:0.0000
## Median : 9.987   Median :75.57   Median :0.0000
## Mean   :11.493   Mean   :68.77   Mean   :0.3212
## 3rd Qu.:14.987   3rd Qu.:89.00   3rd Qu.:1.0000
## Max.   :89.990   Max.   :98.00   Max.   :1.0000
```

```
## NA's :11      NA's :37
```

Use *str* function to provide an overview of data structure for *games*.

```
str(games)
```

```
## 'data.frame':  358 obs. of  7 variables:
## $ appID      : int  280 1250 2300 3340 3960 4770 4870 4890 6200 7670 ...
## $ dateMonth: int  201601 201601 201601 201601 201601 201601 201601 201601 201601 201601 ...
## $ status     : Factor w/ 3 levels "Niche","Popular",...: 3 3 1 1 1 2 1 1 1 3 ...
## $ avgMin      : num   75.5 2103.4 33.7 146.6 98 ...
## $ listPrice: num    9.99 19.99 4.99 4.99 4.99 ...
## $ reviews    : num   83 96 90 10 36.8 ...
## $ multi       : int    0 1 1 0 1 1 0 1 0 0 ...
```

## 2 - Visualization

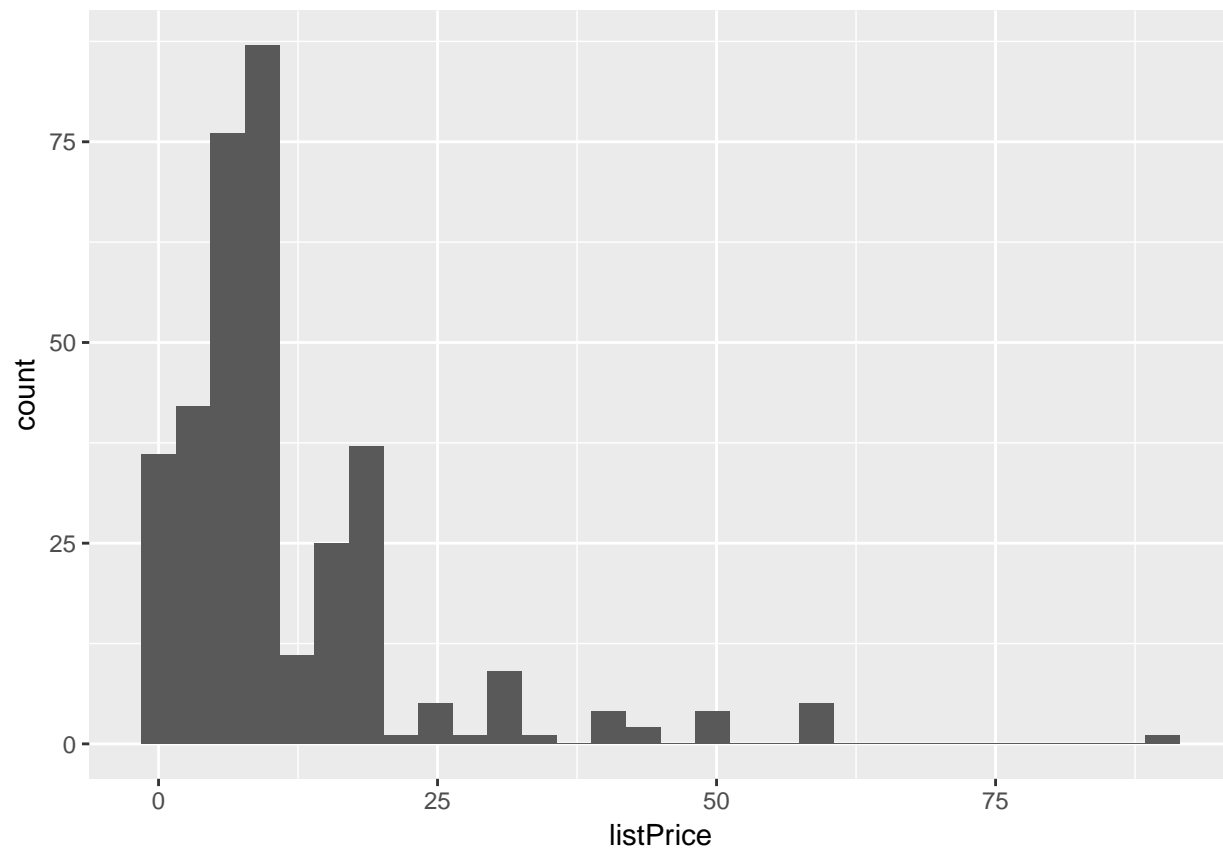
### Histogram

Plot out the distribution of price variable in a histogram.

```
qplot(x = listPrice, data = games, geom = "histogram")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## Warning: Removed 11 rows containing non-finite values (stat_bin).
```

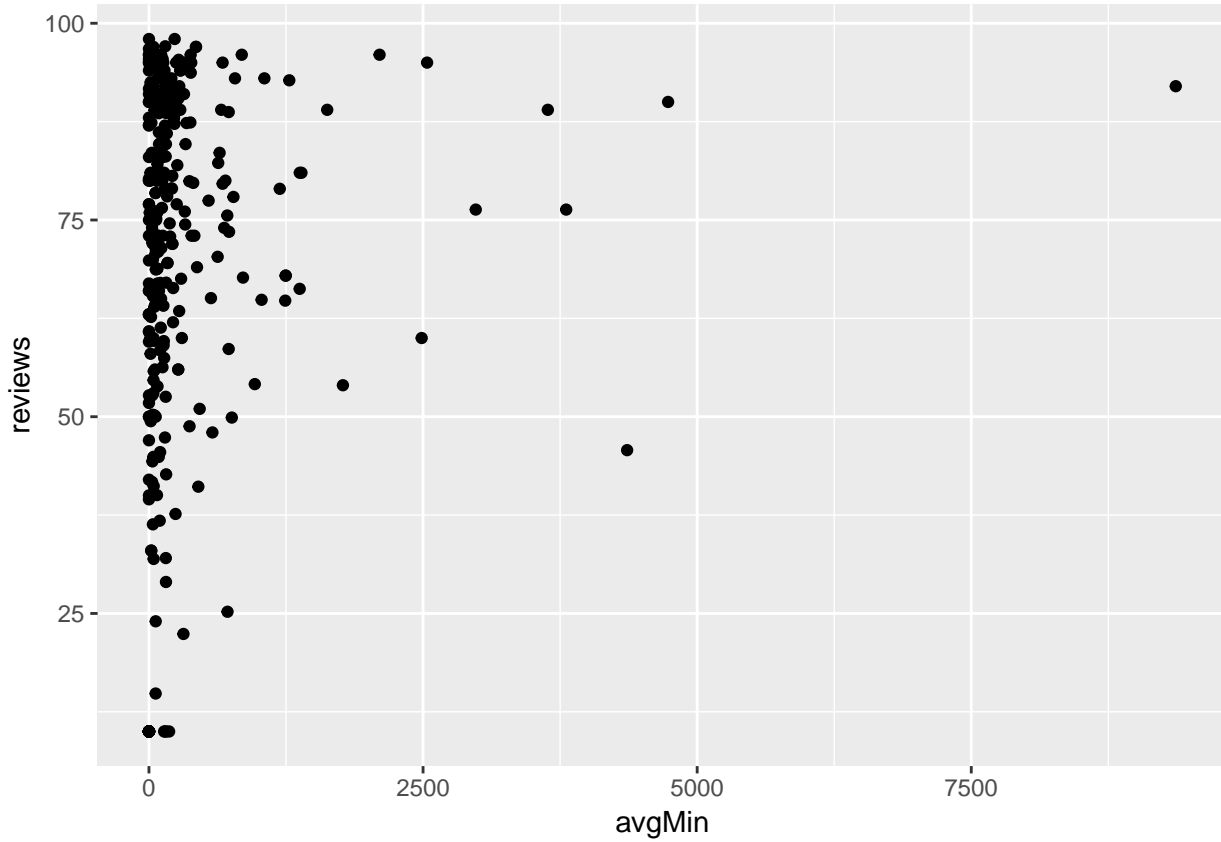


## Scatter Plot

Plot the relation between avgMin (on average how much time players spent in game) and reviews.

```
qplot(x = avgMin, y = reviews, data = games)
```

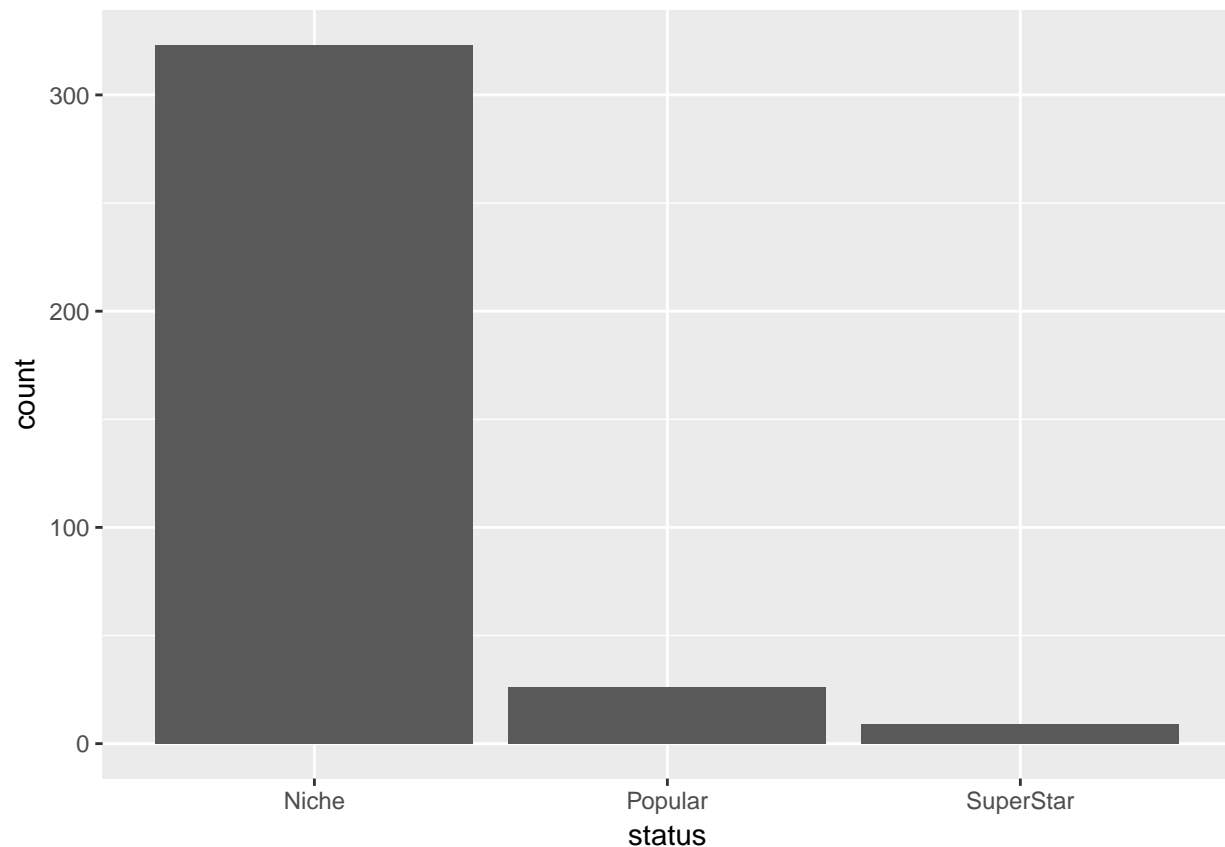
```
## Warning: Removed 37 rows containing missing values (geom_point).
```



## Bar Chart

Plot the frequency of status types in *games*.

```
qplot(x = status, data = games, geom = "bar")
```



### 3 - Subsetting

#### Selecting Rows and Columns

Show only the games that have SuperStar status.

```
games[games$status == "SuperStar",]
```

```
##      appID dateMonth  status    avgMin listPrice reviews multi
## 1      280    201601 SuperStar  75.52385  9.986786      83      0
## 2     1250    201601 SuperStar 2103.36076 19.986786      96      1
## 10     7670    201601 SuperStar  284.50698 19.977143      94      0
## 12     8190    201601 SuperStar  784.62215 14.986786      93      0
## 70    70000    201601 SuperStar  114.52498  9.965357      84      1
## 86   200510    201601 SuperStar 2536.72044 29.977143      95      1
## 130  226320    201601 SuperStar 1375.69520  0.000000      81      0
## 142  236850    201601 SuperStar 9365.33796 39.986786      92      1
## 319  363970    201601 SuperStar 4734.80276  0.000000      90      0
```

Show only the first three rows.

```
games[,1:3]
```

```
##      appID dateMonth  status
## 1      280    201601 SuperStar
## 2     1250    201601 SuperStar
## 3     2300    201601    Niche
## 4     3340    201601    Niche
```

## 5	3960	201601	Niche
## 6	4770	201601	Popular
## 7	4870	201601	Niche
## 8	4890	201601	Niche
## 9	6200	201601	Niche
## 10	7670	201601	SuperStar
## 11	7970	201601	Niche
## 12	8190	201601	SuperStar
## 13	8310	201601	Niche
## 14	8660	201601	Popular
## 15	9060	201601	Niche
## 16	9850	201601	Niche
## 17	11190	201601	Niche
## 18	11200	201601	Popular
## 19	11360	201601	Niche
## 20	11590	201601	Niche
## 21	12170	201601	Popular
## 22	12440	201601	Niche
## 23	15300	201601	Niche
## 24	15380	201601	Niche
## 25	16020	201601	Niche
## 26	16860	201601	Niche
## 27	17580	201601	Niche
## 28	17700	201601	Popular
## 29	18430	201601	Niche
## 30	20700	201601	Niche
## 31	21910	201601	Niche
## 32	22500	201601	Niche
## 33	23490	201601	Popular
## 34	24400	201601	Niche
## 35	24960	201601	Popular
## 36	29130	201601	Niche
## 37	32420	201601	Niche
## 38	32670	201601	Niche
## 39	32710	201601	Niche
## 40	33440	201601	Niche
## 41	34830	201601	Popular
## 42	35083	201601	Niche
## 43	37240	201601	Niche
## 44	37260	201601	Niche
## 45	38150	201601	Niche
## 46	38460	201601	Niche
## 47	39270	201601	Niche
## 48	40700	201601	Popular
## 49	42200	201601	Niche
## 50	42920	201601	Niche
## 51	45797	201601	Niche
## 52	46250	201601	Niche
## 53	46370	201601	Niche
## 54	47700	201601	Niche
## 55	47894	201601	Niche
## 56	47895	201601	Niche
## 57	49320	201601	Niche
## 58	50300	201601	Popular

## 59	50510	201601	Niche
## 60	55203	201601	Niche
## 61	55343	201601	Niche
## 62	57690	201601	Popular
## 63	58230	201601	Niche
## 64	58570	201601	Niche
## 65	63004	201601	Niche
## 66	63700	201601	Niche
## 67	64014	201601	Niche
## 68	65070	201601	Niche
## 69	65540	201601	Niche
## 70	70000	201601	SuperStar
## 71	70300	201601	Popular
## 72	71160	201601	Niche
## 73	73119	201601	Niche
## 74	80000	201601	Niche
## 75	80310	201601	Niche
## 76	80330	201601	Niche
## 77	90007	201601	Niche
## 78	91330	201601	Niche
## 79	92400	201601	Niche
## 80	93207	201601	Niche
## 81	94300	201601	Niche
## 82	97385	201601	Niche
## 83	99612	201601	Niche
## 84	107600	201601	Niche
## 85	115100	201601	Popular
## 86	200510	201601	SuperStar
## 87	202670	201601	Niche
## 88	204080	201601	Popular
## 89	204282	201601	Niche
## 90	205690	201601	Niche
## 91	206112	201601	Niche
## 92	206125	201601	Niche
## 93	206150	201601	Niche
## 94	207320	201601	Niche
## 95	208110	201601	Niche
## 96	208353	201601	Niche
## 97	208379	201601	Niche
## 98	208470	201601	Niche
## 99	208860	201601	Niche
## 100	209750	201601	Niche
## 101	210042	201601	Niche
## 102	210061	201601	Niche
## 103	210062	201601	Niche
## 104	210083	201601	Niche
## 105	210892	201601	Niche
## 106	212780	201601	Niche
## 107	215249	201601	Niche
## 108	215303	201601	Niche
## 109	215309	201601	Niche
## 110	216413	201601	Niche
## 111	216421	201601	Niche
## 112	217690	201601	Niche

## 113	218180	201601	Niche
## 114	218700	201601	Niche
## 115	219380	201601	Niche
## 116	219950	201601	Niche
## 117	222102	201601	Niche
## 118	222121	201601	Niche
## 119	222137	201601	Niche
## 120	222558	201601	Niche
## 121	222591	201601	Niche
## 122	222609	201601	Niche
## 123	224141	201601	Niche
## 124	224144	201601	Niche
## 125	224380	201601	Niche
## 126	224460	201601	Niche
## 127	224900	201601	Niche
## 128	224920	201601	Niche
## 129	225260	201601	Popular
## 130	226320	201601	SuperStar
## 131	226666	201601	Niche
## 132	227000	201601	Niche
## 133	227020	201601	Niche
## 134	229080	201601	Niche
## 135	229870	201601	Niche
## 136	233410	201601	Niche
## 137	234022	201601	Niche
## 138	234570	201601	Niche
## 139	234710	201601	Popular
## 140	235400	201601	Niche
## 141	235620	201601	Niche
## 142	236850	201601	SuperStar
## 143	237350	201601	Niche
## 144	237621	201601	Niche
## 145	237700	201601	Niche
## 146	238240	201601	Niche
## 147	238240	201601	Niche
## 148	238630	201601	Niche
## 149	239120	201601	Niche
## 150	239200	201601	Popular
## 151	239430	201601	Niche
## 152	240620	201601	Niche
## 153	242300	201601	Niche
## 154	242880	201601	Niche
## 155	242940	201601	Niche
## 156	243040	201601	Niche
## 157	243280	201601	Niche
## 158	244030	201601	Niche
## 159	244060	201601	Niche
## 160	245280	201601	Niche
## 161	246250	201601	Niche
## 162	247730	201601	Popular
## 163	248570	201601	Popular
## 164	249310	201601	Niche
## 165	254260	201601	Niche
## 166	255370	201601	Niche

## 167	255940	201601	Niche
## 168	256521	201601	Niche
## 169	256611	201601	Niche
## 170	257630	201601	Niche
## 171	257690	201601	Niche
## 172	258640	201601	Niche
## 173	260430	201601	Niche
## 174	260430	201601	Niche
## 175	260570	201601	Niche
## 176	260850	201601	Niche
## 177	262210	201601	Niche
## 178	262470	201601	Niche
## 179	262980	201601	Niche
## 180	263100	201601	Niche
## 181	263320	201601	Niche
## 182	263420	201601	Niche
## 183	263760	201601	Niche
## 184	264280	201601	Niche
## 185	264440	201601	Niche
## 186	265146	201601	Niche
## 187	265350	201601	Niche
## 188	265380	201601	Niche
## 189	268360	201601	Niche
## 190	268360	201601	Niche
## 191	269350	201601	Niche
## 192	269670	201601	Niche
## 193	270950	201601	Niche
## 194	271418	201601	Niche
## 195	271470	201601	Niche
## 196	271490	201601	Niche
## 197	271492	201601	Niche
## 198	271570	201601	Niche
## 199	271760	201601	Niche
## 200	271820	201601	Niche
## 201	271900	201601	Niche
## 202	274250	201601	Niche
## 203	277150	201601	Niche
## 204	277210	201601	Niche
## 205	277560	201601	Niche
## 206	279260	201601	Niche
## 207	279540	201601	Niche
## 208	281560	201601	Niche
## 209	282070	201601	Popular
## 210	282900	201601	Niche
## 211	283550	201601	Niche
## 212	284000	201601	Niche
## 213	284040	201601	Niche
## 214	284830	201601	Niche
## 215	286260	201601	Niche
## 216	286540	201601	Niche
## 217	286960	201601	Niche
## 218	287390	201601	Popular
## 219	287780	201601	Niche
## 220	287860	201601	Niche



##	221	288790	201601	Niche
##	222	293059	201601	Niche
##	223	293600	201601	Niche
##	224	294310	201601	Niche
##	225	294330	201601	Niche
##	226	294958	201601	Niche
##	227	295016	201601	Niche
##	228	296240	201601	Niche
##	229	296300	201601	Niche
##	230	296300	201601	Niche
##	231	296930	201601	Niche
##	232	297020	201601	Niche
##	233	297490	201601	Niche
##	234	297810	201601	Niche
##	235	297841	201601	Niche
##	236	298220	201601	Niche
##	237	299070	201601	Niche
##	238	299240	201601	Niche
##	239	299480	201601	Niche
##	240	299660	201601	Niche
##	241	300440	201601	Niche
##	242	301110	201601	Niche
##	243	301113	201601	Niche
##	244	301560	201601	Niche
##	245	302130	201601	Niche
##	246	302210	201601	Niche
##	247	302250	201601	Niche
##	248	302460	201601	Niche
##	249	304170	201601	Niche
##	250	304290	201601	Niche
##	251	305050	201601	Niche
##	252	305150	201601	Niche
##	253	305180	201601	Niche
##	254	306130	201601	Popular
##	255	307270	201601	Niche
##	256	307500	201601	Niche
##	257	307780	201601	Popular
##	258	310450	201601	Niche
##	259	311060	201601	Niche
##	260	311210	201601	Popular
##	261	312090	201601	Niche
##	262	312300	201601	Niche
##	263	312960	201601	Niche
##	264	313010	201601	Niche
##	265	313420	201601	Niche
##	266	313611	201601	Niche
##	267	314950	201601	Niche
##	268	314980	201601	Niche
##	269	315730	201601	Niche
##	270	316040	201601	Niche
##	271	316050	201601	Niche
##	272	316651	201601	Niche
##	273	316930	201601	Niche
##	274	316930	201601	Niche

##	275	316964	201601	Niche
##	276	317400	201601	Popular
##	277	317410	201601	Niche
##	278	317833	201601	Niche
##	279	318260	201601	Niche
##	280	318440	201601	Niche
##	281	320140	201601	Niche
##	282	320430	201601	Niche
##	283	320880	201601	Niche
##	284	321300	201601	Niche
##	285	322600	201601	Niche
##	286	322930	201601	Niche
##	287	323060	201601	Niche
##	288	323580	201601	Niche
##	289	324690	201601	Niche
##	290	325080	201601	Niche
##	291	325990	201601	Niche
##	292	326160	201601	Niche
##	293	326160	201601	Niche
##	294	326760	201601	Niche
##	295	327150	201601	Niche
##	296	327490	201601	Niche
##	297	328070	201601	Niche
##	298	329110	201601	Niche
##	299	331510	201601	Niche
##	300	332070	201601	Niche
##	301	332340	201601	Niche
##	302	334070	201601	Niche
##	303	334270	201601	Niche
##	304	334420	201601	Niche
##	305	335410	201601	Niche
##	306	337070	201601	Niche
##	307	342200	201601	Niche
##	308	347160	201601	Niche
##	309	347160	201601	Niche
##	310	347410	201601	Niche
##	311	347620	201601	Niche
##	312	348450	201601	Niche
##	313	356570	201601	Niche
##	314	357330	201601	Niche
##	315	357480	201601	Niche
##	316	359370	201601	Niche
##	317	360960	201601	Niche
##	318	361990	201601	Niche
##	319	363970	201601	SuperStar
##	320	365560	201601	Niche
##	321	365890	201601	Niche
##	322	367580	201601	Niche
##	323	367780	201601	Niche
##	324	368500	201601	Niche
##	325	369200	201601	Niche
##	326	370240	201601	Niche
##	327	370360	201601	Niche
##	328	370360	201601	Niche

```
## 329 373180      201601      Niche
## 330 373240      201601      Niche
## 331 373400      201601      Niche
## 332 375520      201601      Niche
## 333 377670      201601      Niche
## 334 377870      201601      Niche
## 335 378660      201601      Niche
## 336 384110      201601      Niche
## 337 384300      201601      Niche
## 338 385070      201601      Niche
## 339 385440      201601      Niche
## 340 385710      201601      Niche
## 341 386880      201601      Niche
## 342 386900      201601      Niche
## 343 388410      201601      Popular
## 344 393420      201601      Niche
## 345 399040      201601      Niche
## 346 400160      201601      Niche
## 347 400240      201601      Niche
## 348 400250      201601      Niche
## 349 400580      201601      Niche
## 350 400660      201601      Niche
## 351 402180      201601      Niche
## 352 403900      201601      Niche
## 353 404430      201601      Niche
## 354 405360      201601      Niche
## 355 408120      201601      Niche
## 356 416130      201601      Niche
## 357 421630      201601      Niche
## 358 423880      201601      Niche
```

## Aggregating

Calculate the average price for each status category, save the results into a data.frame called *gameAvgPrice*.

```
# Using formula interface for aggregate
gameAvgPrice <- aggregate(listPrice ~ status, games, mean)
```

Change the name of the column listPrice in *gameAvgPrice* to avgCategoryPrice.

```
colnames(gameAvgPrice) <- c("status", "avgCategoryPrice")
```

## Combine Datasets

Combine *games* and *gameAvgPrice* so that the average price for category is displayed next to each game. Save the results into *games* data.frame.

I want column avgCategoryPrice to be added to games, in such a way that it shows the average price for the correct status. A game that is a niche player will show 10.87 in avgCategoryPrice for example.

```
games <- merge(games, gameAvgPrice, by="status")
```

## Transformations

Add a new column to *games* named `logListPrice`, and store log transformed list price in this column. Don't forget log of 0 is meaningless, take the necessary precautions to ensure this does not lead to loss of data.

```
games$logListPrice <- log(games$listPrice + 1)
```

## Missing Data

Show how many rows there are in *games*.

```
nrow(games)
```

```
## [1] 358
```

Exclude any observations that are missing data on reviews column. Save the resulting data.frame in *games*

```
games <- games[!is.na(games$reviews),]
```

Show how many rows there are after removal of observations.

```
nrow(games)
```

```
## [1] 321
```

## Dimension Reduction

WIP