

Data Programming

Homework 2

Lucas Lorenzo Jakin

2025-04-01

Solutions

Submit a **.html** file created using Quarto via e-classroom. A sample is attached. Display all code (packages,input) and output.

Import the csv file from <https://www.kaggle.com/datasets/vishnuvarthanrao/windows-store> (login is required). Parse the Date column into a date.

Careful with dealing with missing/NA data at the end. Some functions like sum, min, max have a option. Answer these questions independently:

1. Display the dataframe as a tibble.
2. How long has the data been collected?
3. Among the highest rated apps return the 10 apps (rows) with the highest amount of reviews. Display only the first 3 columns.
4. Show the number of apps created per month.
5. Sort the rows of the dataframe from the least wordy to most wordy app name. Display the first 15.

```
# solve 1-5 here
library(tidyverse)
#importing csv file
df <- read.csv("msft.csv")
as.tibble(df)
```

```
# A tibble: 5,322 x 6
```

Name	Rating	No.of.people.Rated	Category	Date	Price
<chr>	<dbl>	<int>	<chr>	<chr>	<chr>
1 Dynamic Reader	3.5	268	Books	07-0~	Free
2 Chemistry, Organic Chemistry ~	3	627	Books	08-0~	Free

```

3 BookViewer 3.5 593 Books 29-0~ Free
4 Brick Instructions 3.5 684 Books 30-0~ Free
5 Introduction to Python Progra~ 2 634 Books 30-0~ Free
6 Gurbani Reader 4.5 399 Books 18-0~ Free
7 NFO Viewer 4.5 566 Books 13-1~ Free
8 Text to Speech TTS 3 175 Books 17-1~ Free
9 ACK Comics 4 460 Books 20-1~ Free
10 Learn Biology and Human Body ~ 3.5 363 Books 25-1~ Free
# i 5,312 more rows

```

```

#1 Display the dataframe as a tibble
# Parse Date column into a date
df$Date <- dmy(df$Date)
as.tibble(df)

```

```
# A tibble: 5,322 x 6
```

Name <chr>	Rating <dbl>	No.of.people.Rated <int>	Category <chr>	Date <date>	Price <chr>
1 Dynamic Reader	3.5	268	Books	2014-01-07	Free
2 Chemistry, Organic Chemi~	3	627	Books	2014-01-08	Free
3 BookViewer	3.5	593	Books	2016-02-29	Free
4 Brick Instructions	3.5	684	Books	2018-01-30	Free
5 Introduction to Python P~	2	634	Books	2018-01-30	Free
6 Gurbani Reader	4.5	399	Books	2017-01-18	Free
7 NFO Viewer	4.5	566	Books	2012-11-13	Free
8 Text to Speech TTS	3	175	Books	2013-10-17	Free
9 ACK Comics	4	460	Books	2012-12-20	Free
10 Learn Biology and Human ~	3.5	363	Books	2013-11-25	Free

```

# i 5,312 more rows

```

```

#2 How long has the data been collected?
diff_date <- max(df$Date, na.rm=T) - min(df$Date, na.rm=T)
print(diff_date)

```

Time difference of 3529 days

```

#3 10 rows with highest amount of reviews. Display only first 3 columns
highest Rated <- df[df$Rating==5,]
highest Rated <- highest Rated[order(highest Rated$No.of.people.Rated, decreasing = T),]

```

```
highest_rated <- highest_rated[1:10, 1:3]
highest_rated
```

	Name	Rating	No.of.people.Rated
1884	ILN	5	999
320	TookBook	5	998
3407	Contemplatio - Rosary with images	5	998
3242	Five Good Things	5	996
2663	Magenta Health	5	993
1470	BIOSToolKit	5	989
253	Ask Sai	5	988
2039	Parents Worship Day	5	988
5306	Littledot	5	987
4654	See Tow Bus	5	986

```
#4 Show number of apps created per month
appsPerMonth <- table(format(df$Date, "%m"), useNA="always")
appsPerMonth
```

01	02	03	04	05	06	07	08	09	10	11	12	<NA>
559	480	479	381	456	381	422	450	395	512	423	383	1

```
#df %>% mutate(month = format(df$Date, "%m")) %>% count(month) %>% unname()
```

```
#5 Sort rows of dataframe from least wordy to most wordy app name.
```

```
# Displaye first 15
```

```
library(dplyr)
```

```
df %>% arrange(nchar(Name)) %>% head(15)
```

	Name	Rating	No.of.people.Rated	Category	Date	Price
1	NA		948			<NA>
2		5.0	439	Books	2016-04-01	Free
3	T3	4.0	970	Books	2014-03-10	Free
4		2.5	902	Social	2016-03-01	Free
5	TPi	5.0	100	Business	2017-02-03	Free
6	QMR	5.0	505	Developer Tools	2017-01-14	Free
7	RXT	3.5	820	Developer Tools	2015-09-22	Free
8	ICQ	3.5	733	Social	2012-10-12	Free

9	ILN	5.0	999	Social	2019-12-05	Free
10	Tok	3.5	833	Social	2017-04-03	Free
11	RTI	4.5	565	Government and Politics	2013-12-19	Free
12	4.5		511	Lifestyle	2015-10-04	Free
13	N10	3.5	451	Lifestyle	2015-01-30	Free
14	5.0		346	Lifestyle	2016-03-18	Free
15	SHX	5.0	660	Lifestyle	2019-06-17	Free