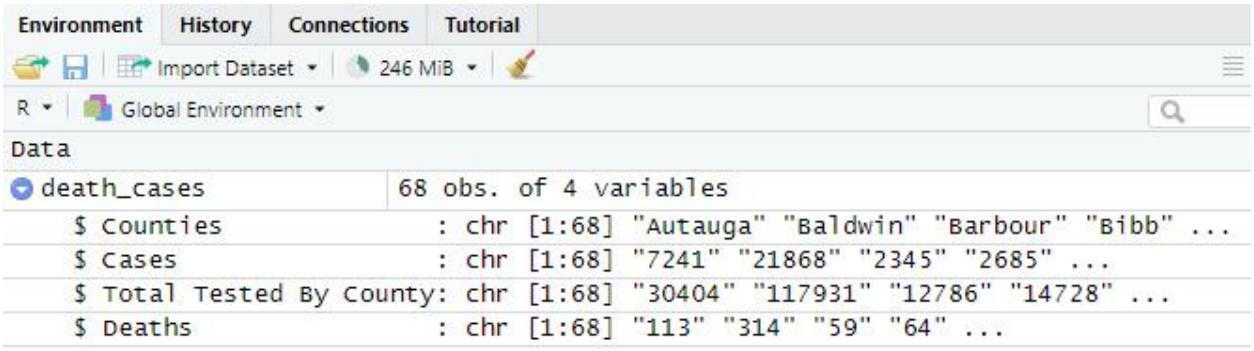
Question1:

Code explanation:

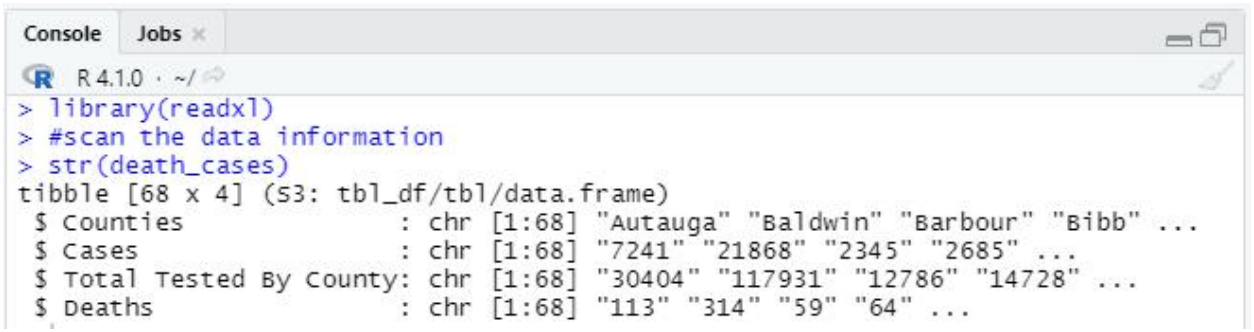
I first install “readxl” and “xlsx” packages, and then use library to call them. library(readxl); library(xlsx)

And then I read the death_cases dataset, (note that here it is lowercase d), and I read the excel file saved in my computer, location of the excel file is: C:/Users/USER/Downloads/COVID-19 in Alabama.xlsx.

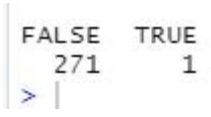
Code: death_cases<-read_xlsx("C:/Users/USER/Downloads/COVID-19 in Alabama.xlsx"). now in the enviroment area (top right of the R studio), you can see the death_cases appears at here. If you click the drop down arrow, you can see what variables in this dataset named “death_cases”:



Now I want to see the data type of each variable: I use code: str(death_cases). Run it I get:



and then I use table(is.na(death_cases)) to find the missing value, result is 271 False and 1 True, meaning we have 271 good values and only one missing values,



so I use is.na(death_cases)to see who is that N/A value. If you run this code, you can see row 68th is N/A, because county in the code counts from 1 and excel counts from 2, so at the end, the 69th line of excel is blank, and the 68th line of the code is a missing value.

```
[65,]    FALSE FALSE                  FALSE  FALSE
[66,]    FALSE FALSE                  FALSE  FALSE
[67,]    FALSE FALSE                  FALSE  FALSE
[68,]     TRUE FALSE                  FALSE  FALSE
```

And then I want to omit the only one missing value, and name the new dataset as “Death_cases”(capital D). Death_case<-na.omit(death_cases). You can see in the environment area, the Death_cases appears there.

```
death_cases                    68 obs. of 4 variables
    $ Counties                 : chr [1:68] "Autauga" "Baldwin" "Barbour" "Bibb" ...
    $ Cases                    : chr [1:68] "7241" "21868" "2345" "2685" ...
    $ Total Tested By County: chr [1:68] "30404" "117931" "12786" "14728" ...
    $ Deaths                   : chr [1:68] "113" "314" "59" "64" ...
```

from now on, the dataset name changes from “death_cases” to “Death_case”

Since I want to calculate Death_rate, I need to convert the datatype,

Death_case$Cases<-as.numeric(Death_case$Cases)

Death_case$`Total Tested By County`<-as.numeric(Death_case$`Total Tested By County`)

Death_case$Deaths<-as.numeric(Death_case$Deaths)

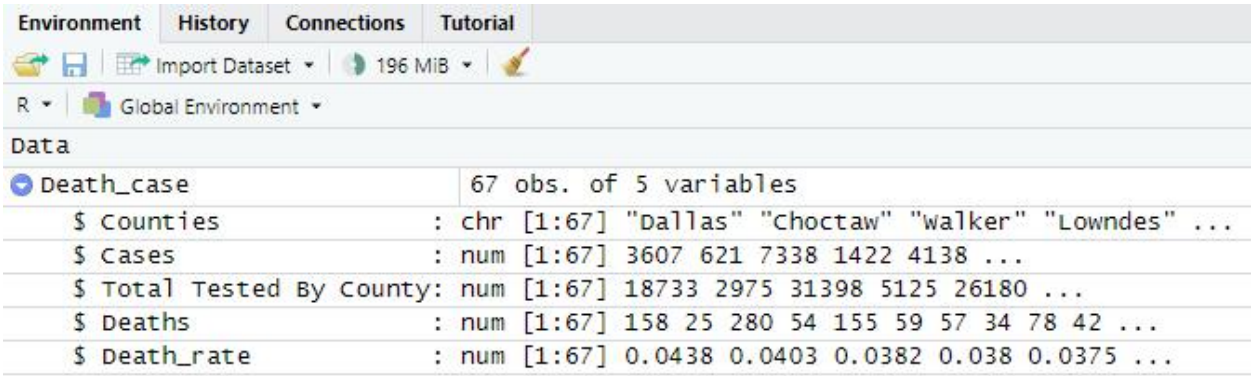
str(Death_case)

Run them, I get the result:

```
> Death_case$Cases<-as.numeric(Death_case$Cases)
> Death_case$`Total Tested By County`<-as.numeric(Death_case$`Total Tested By County
`)
> Death_case$Deaths<-as.numeric(Death_case$Deaths)
> str(Death_case)
tibble [67 x 4] (S3: tbl_df/tbl/data.frame)
 $ Counties              : chr [1:67] "Autauga" "Baldwin" "Barbour" "Bibb" ...
 $ Cases                 : num [1:67] 7241 21868 2345 2685 6945 ...
 $ Total Tested By County: num [1:67] 30404 117931 12786 14728 27790 ...
 $ Deaths                : num [1:67] 113 314 59 64 139 42 71 327 124 45 ...
 - attr(*, "na.action")= 'omit' Named int 68
  ..- attr(*, "names")= chr "68"
```

As you can see, the data type has been changed.

And then I use Death_case$Death_rate<-Death_case$Deaths/Death_case$Cases to calculate Death_rate, and put it to the new vector named: Death_case$Death_rate. Note that the Death_rate is in the dataset named"Death_case". Now the Death_case increase a variable from 4 variables to 5 variables.

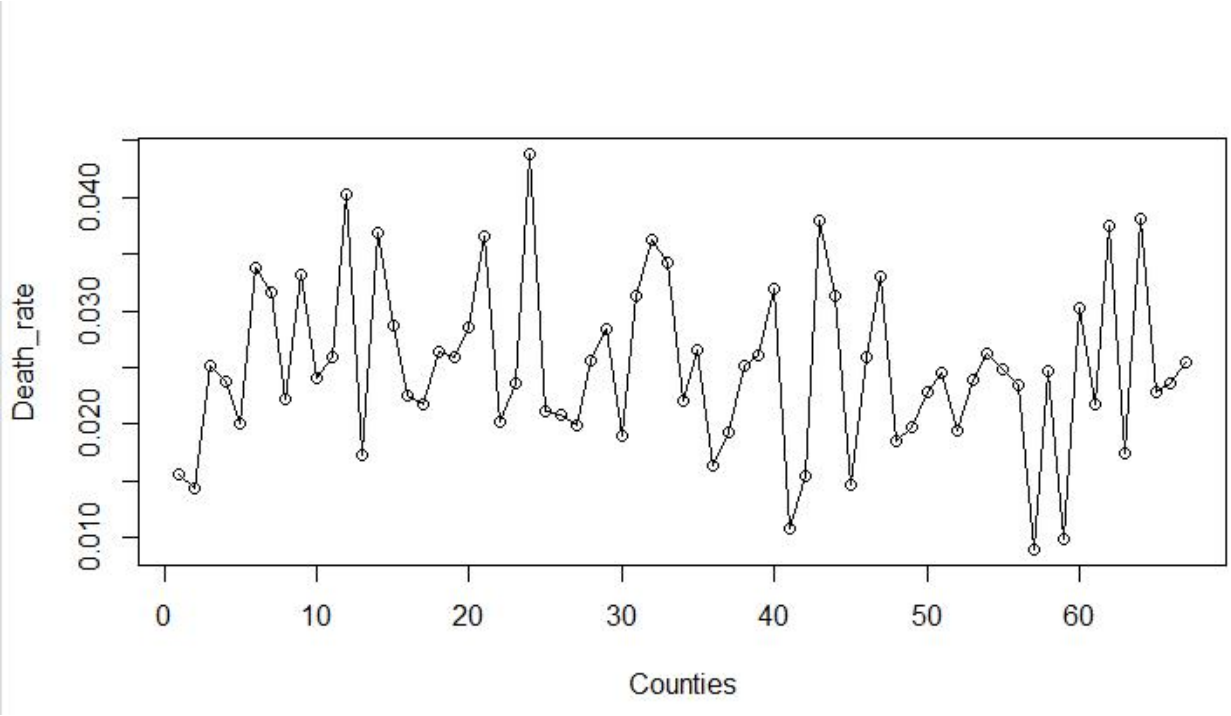


We have one variable added, which is the last one: $Death_rate.

Now you can see the Death_rate table, you click the right side of R studio, which is Environment section, you click the "Death_case", (capital D). you will get the table as follows:

Counties	Cases	Total Tested By County	Deaths	Death_rate
Autauga	7241	30404	113	0.015605579
Baldwin	21868	117931	314	0.014358881
Barbour	2345	12786	59	0.025159915
Bibb	2685	14728	64	0.023836127
Blount	6945	27790	139	0.020014399
Bullock	1243	6239	42	0.033789220
Butler	2242	9135	71	0.031668153
Calhoun	14719	74731	327	0.022216183
Chambers	3728	18911	124	0.033261803
Cherokee	1873	9372	45	0.024025627
Chilton	4476	20225	116	0.025915996
Choctaw	621	2975	25	0.040257649
Clarke	3530	15893	61	0.017280453
Clay	1600	7733	59	0.036875000
Cleburne	1534	6502	44	0.028683181
Coffee	5635	28952	127	0.022537711
Colbert	6413	28993	140	0.021830656
Conecuh	1136	5108	30	0.026408451
Coosa	1117	4087	29	0.025962399
Covington	4275	21281	122	0.028538012
Crenshaw	1559	6711	57	0.036561899

But, As of here, our code is not finished, it just shows us the result of death_rate, I need to sort them by decreasing order, but don't hurry, I want to plot the line chart for the death_rate first. So I use plot(Death_case$Death_rate,type="o",xlab = "Counties",ylab = "Death_rate") to plot the line chart, in this code, $ means the Death_rate is one variable in the dataset Death_case, and "o"means I want the line chart, and xlab is the name of the independent variable, ylab is the name of the dependent variable. And then you can run it and get the line chart as follows:



You may wonder what is the 0 - 68 means, each of them means one county, for example, 2 means Autauga, 3 means Baldwin, 4 means Barbour, etc. Furthermore, from the graph you can see, the death_rate is from 0.010 to 0.4380, each county match their own death_rate.

Ok, now I want to sort the death_rate by decreasing order. The code is : Death_case<-Death_case[order(Death_case$Death_rate, decreasing= T),] Finally, run this code, click the Death_case(remember it is a name with Capital D) in the environment (top right side of R studio) area. You will see the table as follows, the death_rate has been sorted by decreasing order :

	Counties	Cases	Total Tested By County	Deaths	Death_rate
1	Dallas	3607	18733	158	0.043803715
2	Choctaw	621	2975	25	0.040257649
3	Walker	7338	31398	280	0.038157536
4	Lowndes	1422	5125	54	0.037974684
5	Tallapoosa	4138	26180	155	0.037457709
6	Clay	1600	7733	59	0.036875000
7	Crenshaw	1559	6711	57	0.036561899
8	Greene	936	4678	34	0.036324786
9	Hale	2272	10748	78	0.034330986
10	Bullock	1243	6239	42	0.033789220
11	Chambers	3728	18911	124	0.033261803
12	Marion	3240	13388	107	0.033024691
13	Lawrence	3133	10849	100	0.031918289
14	Butler	2242	9135	71	0.031668153
15	Geneva	2579	13693	81	0.031407522
16	Macon	1630	9216	51	0.031288344
17	Sumter	1057	5416	32	0.030274361
18	Cleburne	1534	6502	44	0.028683181
19	Covington	4275	21281	122	0.028538012
20	Fayette	2188	9274	62	0.028336380
21	Houston	10781	63808	287	0.026620907
22	Conecuh	1136	5108	30	0.026408451
23	Pickens	2369	11667	62	0.026171380

Showing 1 to 23 of 67 entries, 5 total columns

	Counties	Cases	Total Tested By County	Deaths	Death_rate
1	Dallas	3607	18733	158	0.043803715
2	Choctaw	621	2975	25	0.040257649
3	Walker	7338	31398	280	0.038157536
4	Lowndes	1422	5125	54	0.037974684
5	Tallapoosa	4138	26180	155	0.037457709
6	Clay	1600	7733	59	0.036875000
7	Crenshaw	1559	6711	57	0.036561899
8	Greene	936	4678	34	0.036324786
9	Hale	2272	10748	78	0.034330986
10	Bullock	1243	6239	42	0.033789220
11	Chambers	3728	18911	124	0.033261803
12	Marion	3240	13388	107	0.033024691
13	Lawrence	3133	10849	100	0.031918289
14	Butler	2242	9135	71	0.031668153
15	Geneva	2579	13693	81	0.031407522
16	Macon	1630	9216	51	0.031288344
17	Sumter	1057	5416	32	0.030274361
18	Cleburne	1534	6502	44	0.028683181
19	Covington	4275	21281	122	0.028538012
20	Fayette	2188	9274	62	0.028336380

	Counties	Cases	Total Tested By County	Deaths	Death_rate
19	Covington	4275	21281	122	0.028538012
20	Fayette	2188	9274	62	0.028336380
21	Houston	10781	63808	287	0.026620907
22	Conecuh	1136	5108	30	0.026408451
23	Pickens	2369	11667	62	0.026171380
24	Lauderdale	9603	45401	250	0.026033531
25	Coosa	1117	4087	29	0.025962399
26	Marengo	2505	13168	65	0.025948104
27	Chilton	4476	20225	116	0.025915996
28	Etowah	14175	64239	364	0.025679012
29	Winston	2834	12118	72	0.025405787
30	Lamar	1468	6575	37	0.025204360
31	Barbour	2345	12786	59	0.025159915
32	Pike	3143	16078	78	0.024817054
33	St. Clair	10162	48252	251	0.024699862
34	Montgomery	25081	119591	614	0.024480683
35	Cherokee	1873	9372	45	0.024025627
36	Perry	1086	5103	26	0.023941068
37	Bibb	2685	14728	64	0.023836127
38	Wilcox	1269	6161	30	0.023640662
39	Dale	4928	28489	116	0.023538961
40	Randolph	1875	9769	44	0.023466667
41	Washington	1705	7696	39	0.022873900

Showing 19 to 41 of 67 entries, 5 total columns

	Counties	Cases	Total Tested By County	Deaths	Death_rate
22	Conecuh	1136	5108	30	0.026408451
23	Pickens	2369	11667	62	0.026171380
24	Lauderdale	9603	45401	250	0.026033531
25	Coosa	1117	4087	29	0.025962399
26	Marengo	2505	13168	65	0.025948104
27	Chilton	4476	20225	116	0.025915996
28	Etowah	14175	64239	364	0.025679012
29	Winston	2834	12118	72	0.025405787
30	Lamar	1468	6575	37	0.025204360
31	Barbour	2345	12786	59	0.025159915
32	Pike	3143	16078	78	0.024817054
33	St. Clair	10162	48252	251	0.024699862
34	Montgomery	25081	119591	614	0.024480683
35	Cherokee	1873	9372	45	0.024025627
36	Perry	1086	5103	26	0.023941068
37	Bibb	2685	14728	64	0.023836127
38	Wilcox	1269	6161	30	0.023640662
39	Dale	4928	28489	116	0.023538961
40	Randolph	1875	9769	44	0.023466667
41	Washington	1705	7696	39	0.022873900
42	Monroe	1800	9616	41	0.022777778
43	Coffee	5635	28952	127	0.022537711

Showing 21 to 44 of 67 entries, 5 total columns

	Counties	Cases	Total Tested By County	Deaths	Death_rate
40	Randolph	1875	9769	44	0.023466667
41	Washington	1705	7696	39	0.022873900
42	Monroe	1800	9616	41	0.022777778
43	Coffee	5635	28952	127	0.022537711
44	Calhoun	14719	74731	327	0.022216183
45	Henry	1945	9467	43	0.022107969
46	Colbert	6413	28993	140	0.021830656
47	Talladega	8460	40210	184	0.021749409
48	DeKalb	8972	32886	190	0.021176995
49	Elmore	10293	54336	214	0.020790829
50	Cullman	9952	45311	201	0.020196945
51	Blount	6945	27790	139	0.020014399
52	Escambia	4016	16848	80	0.019920319
53	Mobile	42105	213507	831	0.019736373
54	Morgan	14629	58909	285	0.019481851
55	Jefferson	81003	457504	1566	0.019332617
56	Franklin	4313	16599	82	0.019012288
57	Marshall	12453	49320	230	0.018469445
58	Tuscaloosa	26173	164765	458	0.017498949
59	Clarke	3530	15893	61	0.017280453
60	Jackson	6932	32486	113	0.016301212
61	Autauga	7241	30404	113	0.015605579
62	Limestone	10179	41000	157	0.015423912

Showing 40 to 62 of 67 entries, 5 total columns

Filter

	Counties	Cases	Total Tested By County	Deaths	Death_rate
47	Talladega	8460	40210	184	0.021749409
48	DeKalb	8972	32886	190	0.021176995
49	Elmore	10293	54336	214	0.020790829
50	Cullman	9952	45311	201	0.020196945
51	Blount	6945	27790	139	0.020014399
52	Escambia	4016	16848	80	0.019920319
53	Mobile	42105	213507	831	0.019736373
54	Morgan	14629	58909	285	0.019481851
55	Jefferson	81003	457504	1566	0.019332617
56	Franklin	4313	16599	82	0.019012288
57	Marshall	12453	49320	230	0.018469445
58	Tuscaloosa	26173	164765	458	0.017498949
59	Clarke	3530	15893	61	0.017280453
60	Jackson	6932	32486	113	0.016301212
61	Autauga	7241	30404	113	0.015605579
62	Limestone	10179	41000	157	0.015423912
63	Madison	35690	221616	525	0.014710003
64	Baldwin	21868	117931	314	0.014358881
65	Lee	16278	95075	176	0.010812139
66	Shelby	25607	134936	254	0.009919163
67	Russell	4548	21250	41	0.009014952

Showing 47 to 67 of 67 entries, 5 total columns

Then I want to do the summary statistics: summary(Death_case), the result is:

```
> summary(Death_case)
   Counties              Cases        Total Tested By County      Deaths      
 Length:67          Min.   :  621   Min.   :  2975         Min.   :  25.0  
 Class :character   1st Qu.: 1874   1st Qu.:  9323         1st Qu.:  52.5  
 Mode  :character   Median : 3728   Median : 16848         Median : 100.0  
                    Mean   : 8189   Mean   : 41835         Mean   : 168.7  
                    3rd Qu.: 9778   3rd Qu.: 43156         3rd Qu.: 195.5  
                    Max.   :81003   Max.   :457504         Max.   :1566.0  
   Death_rate       
 Min.   :0.009015  
 1st Qu.:0.020106  
 Median :0.024481  
 Mean   :0.025075  
 3rd Qu.:0.029479  
 Max.   :0.043804
```

	Cases	Total Tested by County	Deaths	Death_rate
min	621	2975	25	0.009015
1st Qu	1874	9323	52.5	0.020106
Median	3728	16848	100.0	0.024481
Mean	8189	41835	168.7	0.025075
3rd Qu	9778	43156	195.5	0.029479
Max	81003	457504	1566.0	0.043804

As you can see, the Russell county has the lowest death_rate, 0.9%

Dallas county has the highest death_rate, 4.3%.

Pike county has 2.4% death_rate, under the mean death_rate of all alabama state.

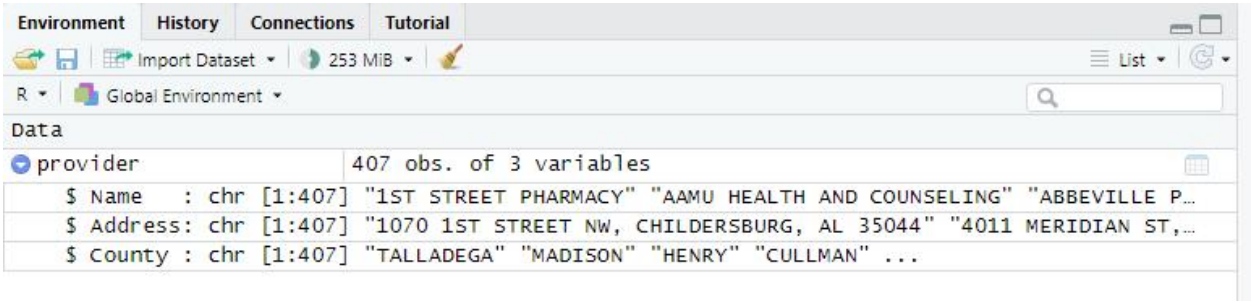
There are 36 counties are under the mean, which is 2.5% death_rate.

Question2 of mid-term:

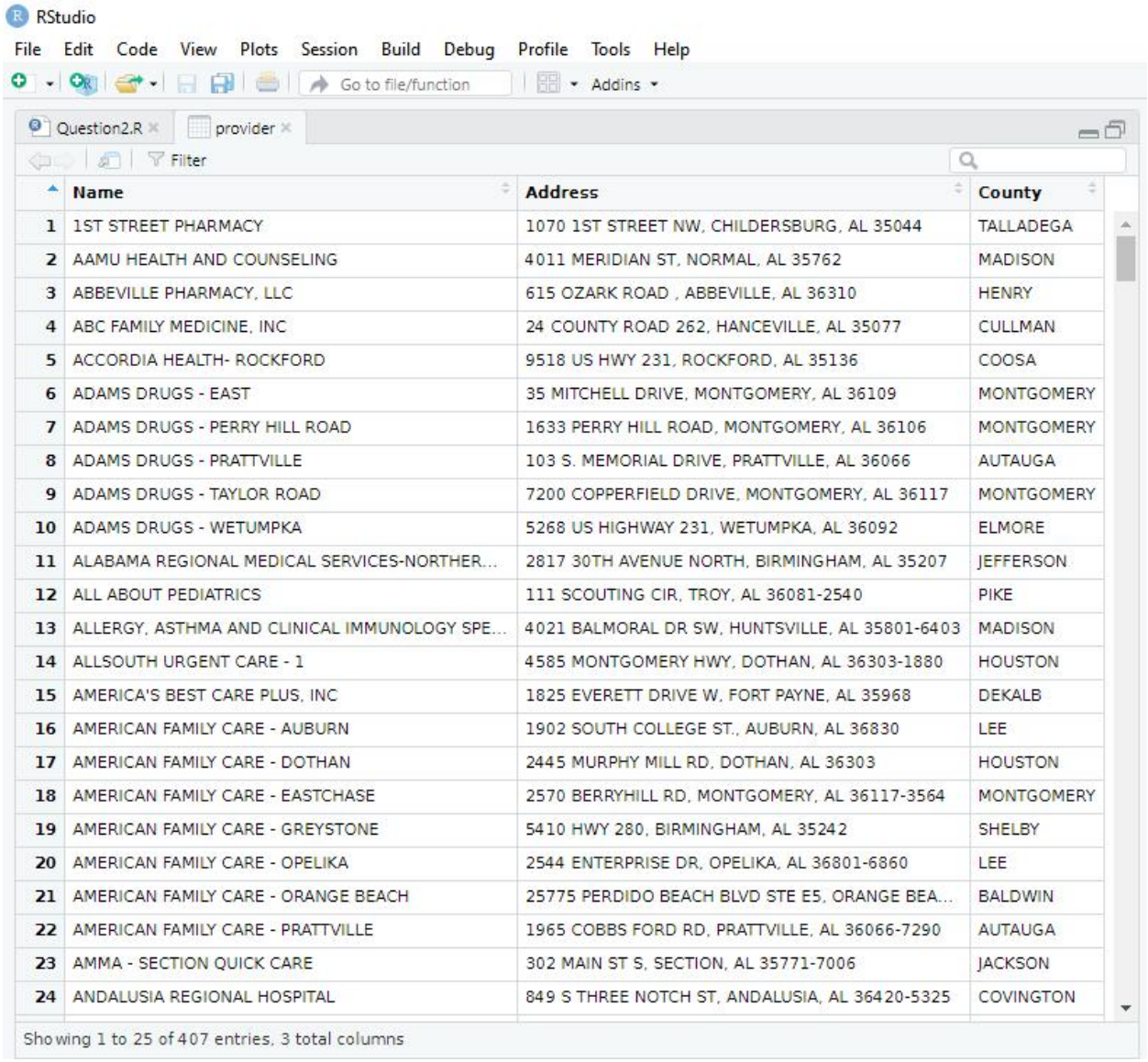
First, I install the package “dplyr”, install.package("dplyr") and install.package(readxl). Or you can use tool to intall it.

Then I use library to recall it. library(dplyr). And then I recall the package readxl: library(readxl).

First, I want to read the excel document saved in my computer and assign it to provider: provider<-read_xlsx("C:/Users/USER/Downloads/COVID-19 in Alabama Vaccine Providers.xlsx"). Now you can see the environment group, the provider has 407 observations of 3 variables.



If you click the “provider” button, the file will appear in R studio like this:



	Name	Address	County
1	1ST STREET PHARMACY	1070 1ST STREET NW, CHILDERSBURG, AL 35044	TALLADEGA
2	AAMU HEALTH AND COUNSELING	4011 MERIDIAN ST, NORMAL, AL 35762	MADISON
3	ABBEVILLE PHARMACY, LLC	615 OZARK ROAD , ABBEVILLE, AL 36310	HENRY
4	ABC FAMILY MEDICINE, INC	24 COUNTY ROAD 262, HANCEVILLE, AL 35077	CULLMAN
5	ACCORDIA HEALTH- ROCKFORD	9518 US HWY 231, ROCKFORD, AL 35136	COOSA
6	ADAMS DRUGS - EAST	35 MITCHELL DRIVE, MONTGOMERY, AL 36109	MONTGOMERY
7	ADAMS DRUGS - PERRY HILL ROAD	1633 PERRY HILL ROAD, MONTGOMERY, AL 36106	MONTGOMERY
8	ADAMS DRUGS - PRATTVILLE	103 S. MEMORIAL DRIVE, PRATTVILLE, AL 36066	AUTAUGA
9	ADAMS DRUGS - TAYLOR ROAD	7200 COPPERFIELD DRIVE, MONTGOMERY, AL 36117	MONTGOMERY
10	ADAMS DRUGS - WETUMPKA	5268 US HIGHWAY 231, WETUMPKA, AL 36092	ELMORE
11	ALABAMA REGIONAL MEDICAL SERVICES-NORTHER...	2817 30TH AVENUE NORTH, BIRMINGHAM, AL 35207	JEFFERSON
12	ALL ABOUT PEDIATRICS	111 SCOUTING CIR, TROY, AL 36081-2540	PIKE
13	ALLERGY, ASTHMA AND CLINICAL IMMUNOLOGY SPE...	4021 BALMORAL DR SW, HUNTSVILLE, AL 35801-6403	MADISON
14	ALLSOUTH URGENT CARE - 1	4585 MONTGOMERY HWY, DOTHAN, AL 36303-1880	HOUSTON
15	AMERICA'S BEST CARE PLUS, INC	1825 EVERETT DRIVE W, FORT PAYNE, AL 35968	DEKALB
16	AMERICAN FAMILY CARE - AUBURN	1902 SOUTH COLLEGE ST., AUBURN, AL 36830	LEE
17	AMERICAN FAMILY CARE - DOTHAN	2445 MURPHY MILL RD, DOTHAN, AL 36303	HOUSTON
18	AMERICAN FAMILY CARE - EASTCHASE	2570 BERRYHILL RD, MONTGOMERY, AL 36117-3564	MONTGOMERY
19	AMERICAN FAMILY CARE - GREYSTONE	5410 HWY 280, BIRMINGHAM, AL 35242	SHELBY
20	AMERICAN FAMILY CARE - OPELIKA	2544 ENTERPRISE DR, OPELIKA, AL 36801-6860	LEE
21	AMERICAN FAMILY CARE - ORANGE BEACH	25775 PERDIDO BEACH BLVD STE E5, ORANGE BEA...	BALDWIN
22	AMERICAN FAMILY CARE - PRATTVILLE	1965 COBBS FORD RD, PRATTVILLE, AL 36066-7290	AUTAUGA
23	AMMA - SECTION QUICK CARE	302 MAIN ST S, SECTION, AL 35771-7006	JACKSON
24	ANDALUSIA REGIONAL HOSPITAL	849 S THREE NOTCH ST, ANDALUSIA, AL 36420-5325	COVINGTON

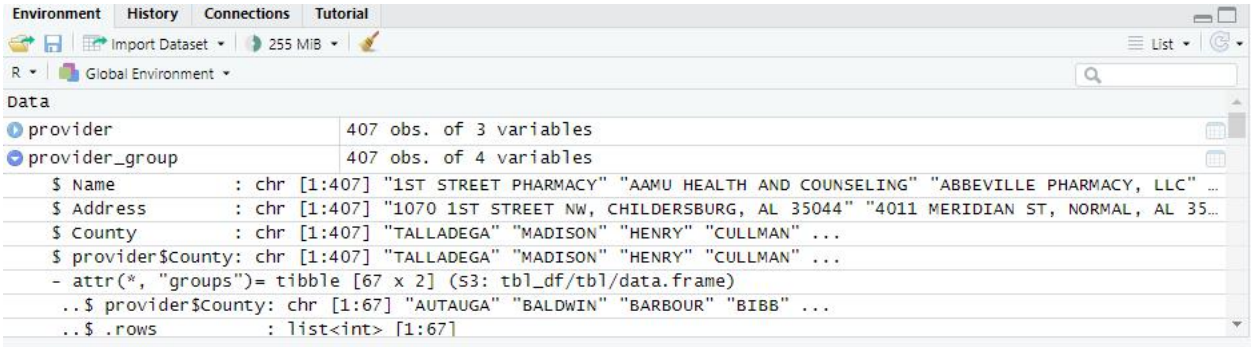
Showing 1 to 25 of 407 entries, 3 total columns

Then I go back to the data source area, I want to know if there is N/A value in the provider table. Code: table(is.na(provider)). If you run it, it returns this:

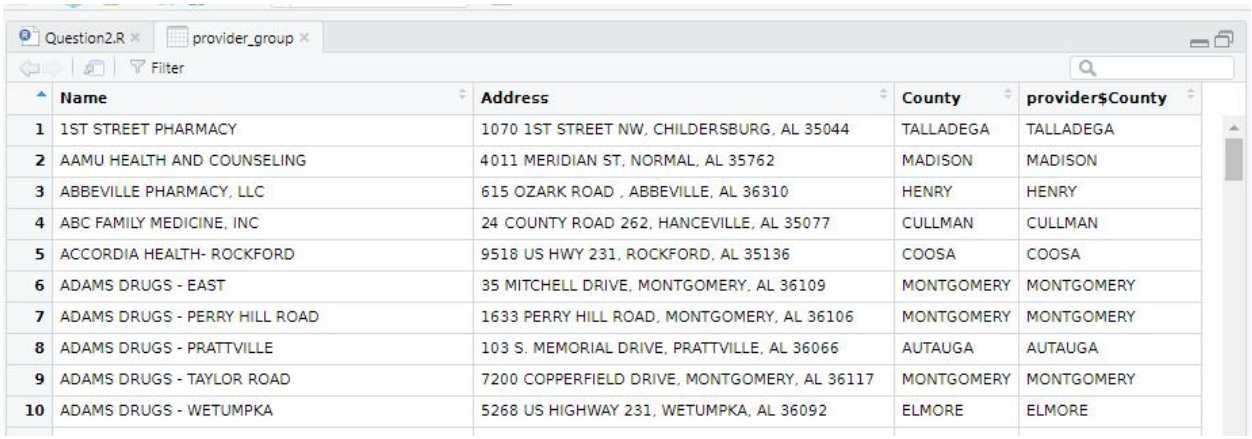
```
FALSE
 1221
```

This means, no N/A value in the total of 1221 values.

Then I want to group these providers by county, code: provider_group<- group_by(provider, provider$County). After you run this code, the group_by variable will appear in the environment area



Then I want to view the table after groupby. Code: View(provider_group). Then the table appears automatically:



	Name	Address	County	provider$County
1	1ST STREET PHARMACY	1070 1ST STREET NW, CHILDERSBURG, AL 35044	TALLADEGA	TALLADEGA
2	AAMU HEALTH AND COUNSELING	4011 MERIDIAN ST, NORMAL, AL 35762	MADISON	MADISON
3	ABBEVILLE PHARMACY, LLC	615 OZARK ROAD , ABBEVILLE, AL 36310	HENRY	HENRY
4	ABC FAMILY MEDICINE, INC	24 COUNTY ROAD 262, HANCEVILLE, AL 35077	CULLMAN	CULLMAN
5	ACCORDIA HEALTH- ROCKFORD	9518 US HWY 231, ROCKFORD, AL 35136	COOSA	COOSA
6	ADAMS DRUGS - EAST	35 MITCHELL DRIVE, MONTGOMERY, AL 36109	MONTGOMERY	MONTGOMERY
7	ADAMS DRUGS - PERRY HILL ROAD	1633 PERRY HILL ROAD, MONTGOMERY, AL 36106	MONTGOMERY	MONTGOMERY
8	ADAMS DRUGS - PRATTVILLE	103 S. MEMORIAL DRIVE, PRATTVILLE, AL 36066	AUTAUGA	AUTAUGA
9	ADAMS DRUGS - TAYLOR ROAD	7200 COPPERFIELD DRIVE, MONTGOMERY, AL 36117	MONTGOMERY	MONTGOMERY
10	ADAMS DRUGS - WETUMPKA	5268 US HIGHWAY 231, WETUMPKA, AL 36092	ELMORE	ELMORE

Next I want to sort the provider$county by ascending order.

provider_group<- provider_group[order(provider_group$`provider$County`,decreasing=F),]

the result is:

Question2.R × provider_group ×

Filter

	Name	Address	County	provider$County
1	ADAMS DRUGS - PRATTVILLE	103 S. MEMORIAL DRIVE, PRATTVILLE, AL 36066	AUTAUGA	AUTAUGA
2	AMERICAN FAMILY CARE - PRATTVILLE	1965 COBBS FORD RD, PRATTVILLE, AL 36066-7290	AUTAUGA	AUTAUGA
3	AUTAUGA CO HLTH DEPT	219 N COURT ST, PRATTVILLE, AL 36067-3003	AUTAUGA	AUTAUGA
4	WYBENGA FAMILY HEALTHCARE	564 N MEMORIAL DR, PRATTVILLE, AL 36067-2132	AUTAUGA	AUTAUGA
5	AMERICAN FAMILY CARE - ORANGE BEACH	25775 PERDIDO BEACH BLVD STE E5, ORANGE BEA...	BALDWIN	BALDWIN
6	BALDWIN CO HLTH DEPT - ROBERTSDALE	23280 GILBERT DRIVE, ROBERTSDALE, AL 36567-03...	BALDWIN	BALDWIN
7	FAIRHOPE PEDIATRICS	19087B GREENO RD, FAIRHOPE, AL 36532-3899	BALDWIN	BALDWIN
8	FPHC - NORTH BALDWIN FAMILY HEALTH CENTER	201 D'OLIVE STREET, BAY MINETTE, AL 36507	BALDWIN	BALDWIN
9	INFIRMARY EMPLOYEE HEALTH - TH	750 MORPHY AVE, FAIRHOPE, AL 36532-1812	BALDWIN	BALDWIN
10	SOUTH BALDWIN REGIONAL MEDICAL CENTER	1613 N MCKENZIE ST, FOLEY, AL 36535-2247	BALDWIN	BALDWIN
11	BARBOUR CO HLTH DEPT - EUFAULA	634 SCHOOL ST, EUFAULA, AL 36027-2430	BARBOUR	BARBOUR
12	MAINSTREET FAMILY CARE - EUFAULA	1026 SOUTH EUFAULA AVE., EUFAULA, AL 36027	BARBOUR	BARBOUR
13	MEDICAL CENTER BARBOUR	820 W WASHINGTON ST, EUFAULA, AL 36027-1822	BARBOUR	BARBOUR
14	SARHA-CLAYTON FAMILY HEALTH CENTER	7 WESTERN BYP, CLAYTON, AL 36016-2013	BARBOUR	BARBOUR

You can see the providers have been grouped by the County. For example, Autauga county, we have four providers: 1. ADAMS DRUGS-PRATTVILLE, 2. AMERICAN FAMILY CARE-PPATTVILLE; 3.AUTAUGA CO HLTH DEPT; 4. WYBENGA FAMILY HEALTHCARE.

In the Autauga, we have 1st-4th providers; In the Baldwin county, we have 5th-10th county; In the Barbour county, we have 11th-14th providers.

Question2.R × provider_group ×

Filter

	Name	Address	County	provider$County
14	SARHA-CLAYTON FAMILY HEALTH CENTER	7 WESTERN BYP, CLAYTON, AL 36016-2013	BARBOUR	BARBOUR
15	BIBB CO HLTH DEPT	281 ALEXANDER AVE, CENTREVILLE, AL 35042-2953	BIBB	BIBB
16	BIBB MEDICAL ASSOCIATES	208 PIERSON AVE, CENTREVILLE, AL 35042-2918	BIBB	BIBB
17	BIBB MEDICAL CENTER PHARMACY	208 PIERSON AVENUE, CENTREVILLE, AL 35042	BIBB	BIBB
18	CAHABA MEDICAL CARE FOUNDATION-CENTREVILLE	405 BELCHER ST, CENTREVILLE, AL 35042-2946	BIBB	BIBB
19	CAHABA MEDICAL CARE, WOODSTOCK	28921 HIGHWAY 5, WOODSTOCK, AL 35188-3613	BIBB	BIBB
20	BLOUNT CO HLTH DEPT - ONEONTA	1001 LINCOLN AVE, ONEONTA, AL 35121-2533	BLOUNT	BLOUNT
21	CLEVELAND PHARMACY	36321 STATE HWY 79 SUITE 2, CLEVELAND, AL 35049	BLOUNT	BLOUNT
22	MAINSTREET FAMILY CARE - ONEONTA	2022 2ND AVE. EAST, ONEONTA, AL 35121	BLOUNT	BLOUNT
23	PREMIER FAMILY CARE BLOUNTSVILLE	68278 MAIN ST , BLOUNTSVILLE, AL 35031	BLOUNT	BLOUNT
24	PRIMARY CARE SERVICES-BLOUNT LLC	101 LEMLEY DR STE A, ONEONTA, AL 35121-2100	BLOUNT	BLOUNT
25	ST. VINCENT'S BLOUNT	150 GILBREATH DR, ONEONTA, AL 35121-2827	BLOUNT	BLOUNT
26	BULLOCK CO HLTH DEPT	674 HICKS INDUSTRIAL BOULEVARD, UNION SPRING...	BULLOCK	BULLOCK
27	BULLOCK COUNTY HOSPITAL	102 CONECUH AVE W, UNION SPRINGS, AL 36089-1...	BULLOCK	BULLOCK
28	BUTLER CO HLTH DEPT - GREENVILLE	350 AIRPORT RD, GREENVILLE, AL 36037-8822	BUTLER	BUTLER
29	LV STABLER PRIMARY CARE LLC	300 N COLLEGE ST, GREENVILLE, AL 36037-2025	BUTLER	BUTLER
30	PRO-MED PC	302 PAUL STABLER DR, GREENVILLE, AL 36037-3128	BUTLER	BUTLER
31	REGIONAL MEDICAL CENTER OF CENTRAL ALABAMA...	29 L V STABLER DR, GREENVILLE, AL 36037-3850	BUTLER	BUTLER

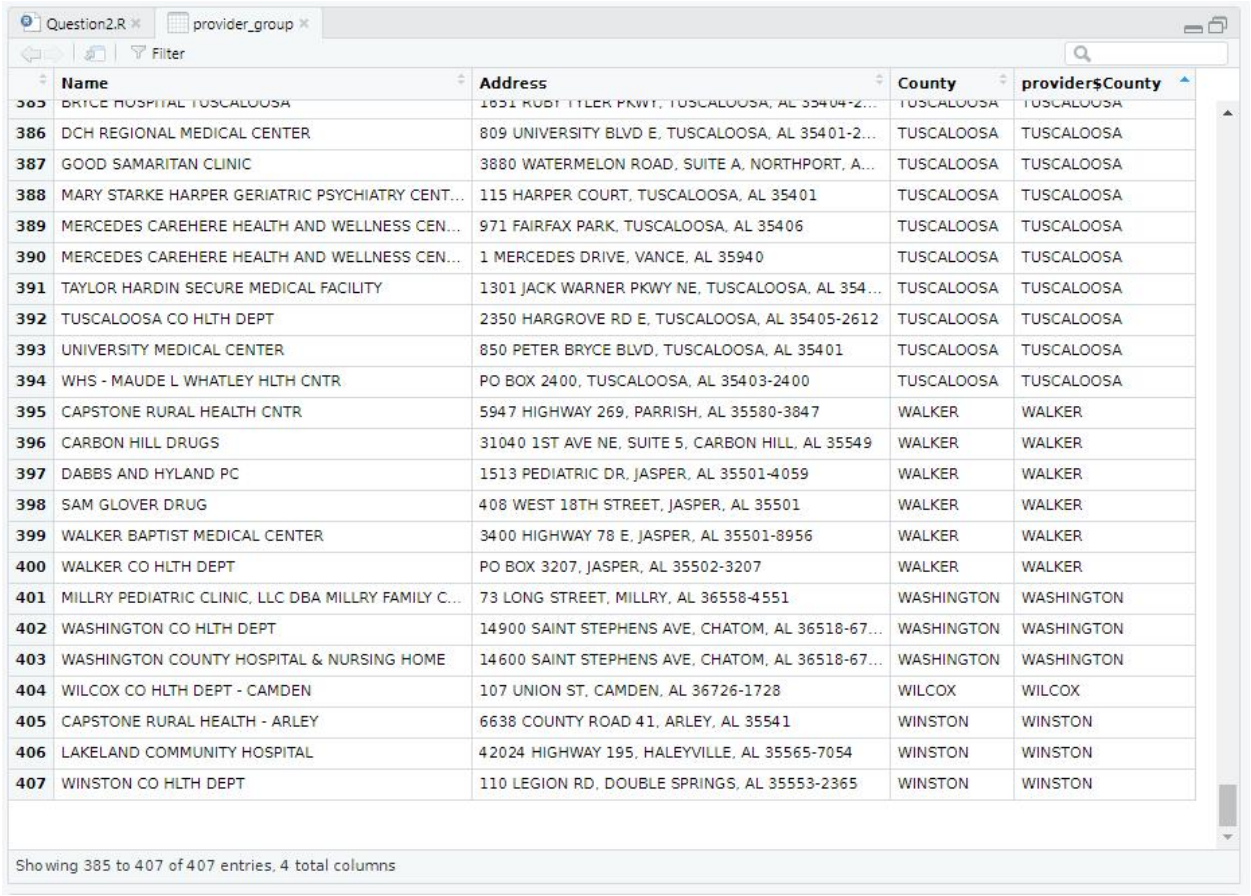
In BIBB county, we have 15th -19th providers; In the BLOUNT County, we have 20th-25th providers; In the BULLOCK County, we have 26th-27th providers; In the BUTLER County, we have 28th – 31th providers;

Question2.R × provider_group ×

Filter

	Name	Address	County	provider$County
30	PRO-MED PC	302 PAUL STABLER DR, GREENVILLE, AL 36037-3128	BUTLER	BUTLER
31	REGIONAL MEDICAL CENTER OF CENTRAL ALABAMA...	29 L V STABLER DR, GREENVILLE, AL 36037-3850	BUTLER	BUTLER
32	ANNISTON MEDICAL CLINIC, PC	1010 CHRISTINE AVE, ANNISTON, AL 36207-5782	CALHOUN	CALHOUN
33	CALHOUN CO HLTH DEPT	3400 MCCLELLAN BLVD, ANNISTON, AL 36201-2128	CALHOUN	CALHOUN
34	HIGHLAND HEALTH SYSTEM	331 EAST 8TH STREET, ANNISTON, AL 36202	CALHOUN	CALHOUN
35	MARTIN'S PHARMACY-OXFORD	610 QUINTARD DRIVE , OXFORD , AL 36203	CALHOUN	CALHOUN
36	NORTHEAST ALABAMA REGIONAL MEDICAL CENTER	400 E 10TH ST, ANNISTON, AL 36207-4716	CALHOUN	CALHOUN
37	PEDIATRIC CARE CTR OF NE AL-DR A MARTIN	304 E 4TH ST, ANNISTON, AL 36207-6012	CALHOUN	CALHOUN
38	SOUTHERN IMMEDIATE CARE	4100 MCCLELLAN BLVD, ANNISTON, AL 36201-2132	CALHOUN	CALHOUN
39	CHAMBERS CO HLTH DEPT - VALLEY	5 MEDICAL PARK N, VALLEY, AL 36854-3607	CHAMBERS	CHAMBERS
40	EAMC-LANIER	4800 48TH ST, VALLEY, AL 36854-3666	CHAMBERS	CHAMBERS
41	FAIRFAX DRUG COMPANY	89 TRAMMELL BLOCK, VALLEY, AL 36854	CHAMBERS	CHAMBERS
42	HOOD'S PHARMACY	4500 20TH AVE, VALLEY, AL 36854	CHAMBERS	CHAMBERS
43	MAINSTREET FAMILY CARE - VALLEY	3306 20TH AVENUE, VALLEY, AL 36854	CHAMBERS	CHAMBERS
44	VALLEY PHARMACY	4103 20TH AVE, VALLEY, AL 36854-3448	CHAMBERS	CHAMBERS
45	CHEROKEE CO HLTH DEPT	833 CEDAR BLUFF RD, CENTRE, AL 35960-1005	CHEROKEE	CHEROKEE
46	CHEROKEE MEDICAL CENTER	400 NORTHWOOD DR, CENTRE, AL 35960-1023	CHEROKEE	CHEROKEE
47	COMPLETE FAMILY CARE	280 INDUSTRIAL BLVD, LEESBURG, AL 35983	CHEROKEE	CHEROKEE
48	DEAN'S DRUGS, INC.	699 CEDAR BLUFF RD., CENTRE, AL 35960	CHEROKEE	CHEROKEE
49	LAKESIDE PRIMARY CARE	1260 CHESNUT BYPASS SUITE A, CENTRE, AL 35960	CHEROKEE	CHEROKEE
50	MAINSTREET FAMILY CARE - CENTRE	1925 WEST MAIN STREET, CENTRE, AL 35960	CHEROKEE	CHEROKEE

In Calhoun County, we have 32th-38th providers; in Chambers County, we have 39th-44th providers; In Cherokee county, we have 45th-50th providers;



Question2.R | provider_group

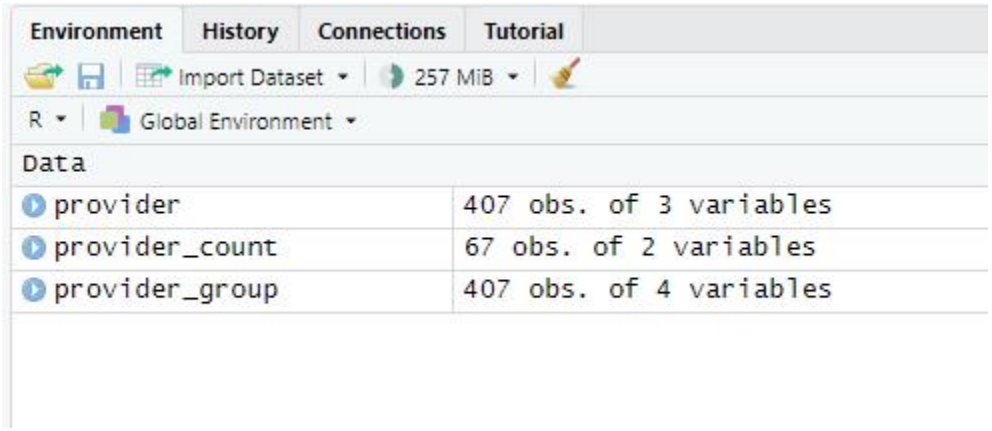
Filter

	Name	Address	County	provider$County
385	BRYCE HOSPITAL TUSCALOOSA	1651 RUBY TYLER PKWY, TUSCALOOSA, AL 35404-2...	TUSCALOOSA	TUSCALOOSA
386	DCH REGIONAL MEDICAL CENTER	809 UNIVERSITY BLVD E, TUSCALOOSA, AL 35401-2...	TUSCALOOSA	TUSCALOOSA
387	GOOD SAMARITAN CLINIC	3880 WATERMELON ROAD, SUITE A, NORTHPORT, A...	TUSCALOOSA	TUSCALOOSA
388	MARY STARKE HARPER GERIATRIC PSYCHIATRY CENT...	115 HARPER COURT, TUSCALOOSA, AL 35401	TUSCALOOSA	TUSCALOOSA
389	MERCEDES CAREHERE HEALTH AND WELLNESS CEN...	971 FAIRFAX PARK, TUSCALOOSA, AL 35406	TUSCALOOSA	TUSCALOOSA
390	MERCEDES CAREHERE HEALTH AND WELLNESS CEN...	1 MERCEDES DRIVE, VANCE, AL 35940	TUSCALOOSA	TUSCALOOSA
391	TAYLOR HARDIN SECURE MEDICAL FACILITY	1301 JACK WARNER PKWY NE, TUSCALOOSA, AL 354...	TUSCALOOSA	TUSCALOOSA
392	TUSCALOOSA CO HLTH DEPT	2350 HARGROVE RD E, TUSCALOOSA, AL 35405-2612	TUSCALOOSA	TUSCALOOSA
393	UNIVERSITY MEDICAL CENTER	850 PETER BRYCE BLVD, TUSCALOOSA, AL 35401	TUSCALOOSA	TUSCALOOSA
394	WHS - MAUDE L WHATLEY HLTH CNTR	PO BOX 2400, TUSCALOOSA, AL 35403-2400	TUSCALOOSA	TUSCALOOSA
395	CAPSTONE RURAL HEALTH CNTR	5947 HIGHWAY 269, PARRISH, AL 35580-3847	WALKER	WALKER
396	CARBON HILL DRUGS	31040 1ST AVE NE, SUITE 5, CARBON HILL, AL 35549	WALKER	WALKER
397	DABBS AND HYLAND PC	1513 PEDIATRIC DR, JASPER, AL 35501-4059	WALKER	WALKER
398	SAM GLOVER DRUG	408 WEST 18TH STREET, JASPER, AL 35501	WALKER	WALKER
399	WALKER BAPTIST MEDICAL CENTER	3400 HIGHWAY 78 E, JASPER, AL 35501-8956	WALKER	WALKER
400	WALKER CO HLTH DEPT	PO BOX 3207, JASPER, AL 35502-3207	WALKER	WALKER
401	MILLRY PEDIATRIC CLINIC, LLC DBA MILLRY FAMILY C...	73 LONG STREET, MILLRY, AL 36558-4551	WASHINGTON	WASHINGTON
402	WASHINGTON CO HLTH DEPT	14900 SAINT STEPHENS AVE, CHATOM, AL 36518-67...	WASHINGTON	WASHINGTON
403	WASHINGTON COUNTY HOSPITAL & NURSING HOME	14600 SAINT STEPHENS AVE, CHATOM, AL 36518-67...	WASHINGTON	WASHINGTON
404	WILCOX CO HLTH DEPT - CAMDEN	107 UNION ST, CAMDEN, AL 36726-1728	WILCOX	WILCOX
405	CAPSTONE RURAL HEALTH - ARLEY	6638 COUNTY ROAD 41, ARLEY, AL 35541	WINSTON	WINSTON
406	LAKELAND COMMUNITY HOSPITAL	42024 HIGHWAY 195, HALEYVILLE, AL 35565-7054	WINSTON	WINSTON
407	WINSTON CO HLTH DEPT	110 LEGION RD, DOUBLE SPRINGS, AL 35553-2365	WINSTON	WINSTON

Showing 385 to 407 of 407 entries, 4 total columns

Until Winston County, we have 405th-407th providers.

Then I want to create a data frame and count the providers by county, and then assign it to a new name: provider_count, using code: provider_count<- summarise(provider_group,count = n()) , now you can see provider_count appears in the environment group,



Then you click the provider_count button:

Question2.R | provider_count

Filter

	provider$County	count
1	AUTAUGA	4
2	BALDWIN	6
3	BARBOUR	4
4	BIBB	5
5	BLOUNT	6
6	BULLOCK	2
7	BUTLER	4
8	CALHOUN	7
9	CHAMBERS	6
10	CHEROKEE	6
11	CHILTON	4
12	CHOCTAW	5
13	CLARKE	7
14	CLAY	6
15	CLEBURNE	2

Showing 1 to 15 of 67 entries, 2 total columns

Then you can see it counts the number of providers in each county.

Then I want to sort the table in decreasing order using code: provider_count<-provider_count[order(provider_count$count,decreasing=T),] and then run it, click the provider_count button in the new environment, you can get the table:


Question2.R × provider_count ×

Filter

	provider$County	count
1	JEFFERSON	28
2	MOBILE	17
3	MONTGOMERY	17
4	MADISON	13
5	MARSHALL	13
6	ETOWAH	12
7	TALLADEGA	11
8	LEE	10
9	TUSCALOOSA	10
10	ELMORE	9
11	HOUSTON	9
12	PIKE	9
13	ST. CLAIR	9
14	DEKALB	8
15	LAUDERDALE	8

Showing 1 to 15 of 67 entries, 2 total columns

Question2.R × provider_count ×

Filter

	provider$County	count
16	CALHOUN	7
17	CLARKE	7
18	JACKSON	7
19	SHELBY	7
20	BALDWIN	6
21	BLOUNT	6
22	CHAMBERS	6
23	CHEROKEE	6
24	CLAY	6
25	DALLAS	6
26	ESCAMBIA	6
27	FRANKLIN	6
28	LIMESTONE	6
29	MONROE	6
30	TALLAPOOSA	6

Showing 16 to 30 of 67 entries, 2 total columns

Question2.R × provider_count ×

Filter

	provider$County	count
29	MONROE	6
30	TALLAPOOSA	6
31	WALKER	6
32	BIBB	5
33	CHOCTAW	5
34	GENEVA	5
35	RANDOLPH	5
36	RUSSELL	5
37	SUMTER	5
38	AUTAUGA	4
39	BARBOUR	4
40	BUTLER	4
41	CHILTON	4
42	COFFEE	4
43	COLBERT	4

Showing 29 to 43 of 67 entries, 2 total columns

Question2.R × provider_count ×

Filter

	provider$County	count
55	LAMAR	3
56	MACON	3
57	MARENGO	3
58	PERRY	3
59	WASHINGTON	3
60	WINSTON	3
61	BULLOCK	2
62	DALE	2
63	GREENE	2
64	LAWRENCE	2
65	LOWNDES	2
66	COOSA	1
67	WILCOX	1

Showing 55 to 67 of 67 entries, 2 total columns

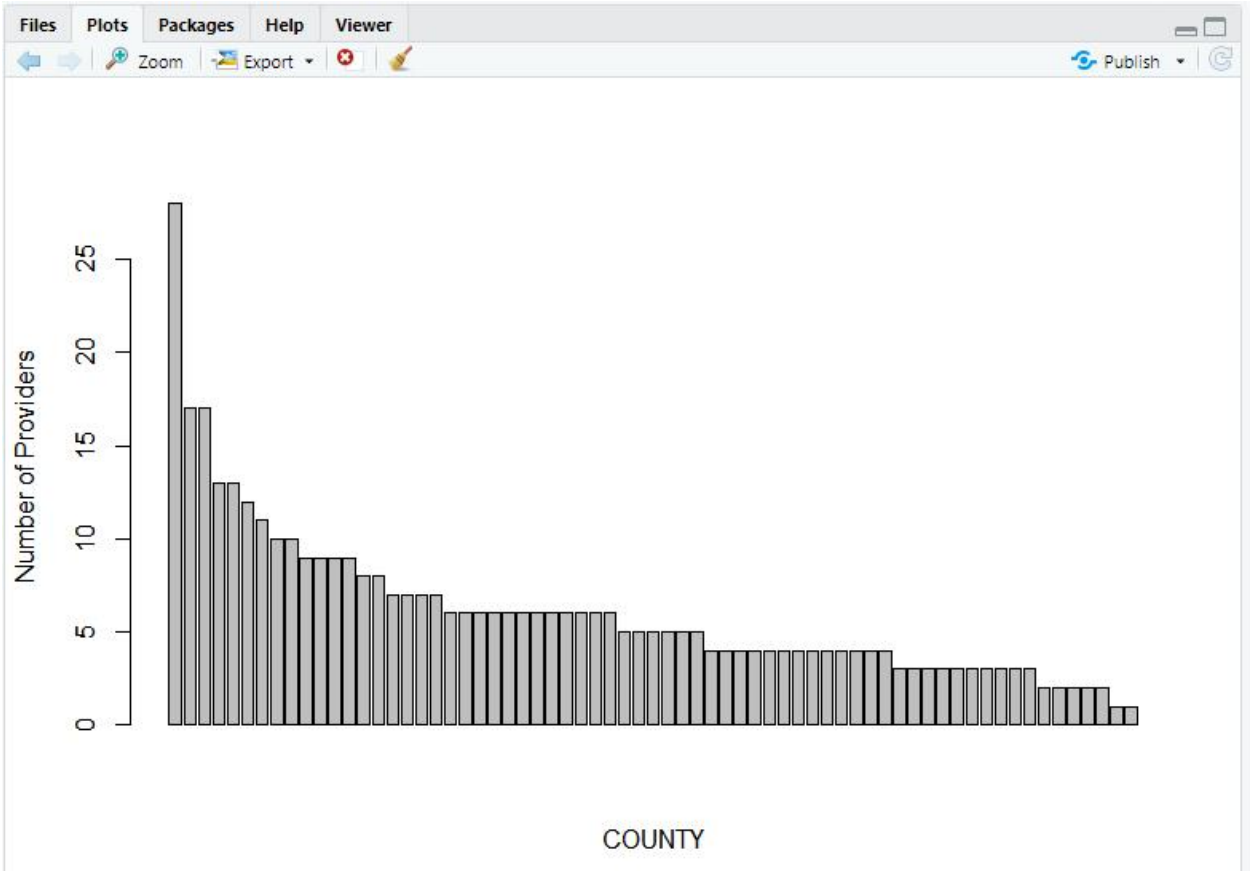
We have 67 counties in total.

So I have finished grouping all the vaccine providers in the same county in AL, and count the number of providers in each county.

Now, I want to do the data visualization, draw a bar chart, x-axis is County, y-axis is number of providers.

```
barplot(provider_count$count, names.arg = NULL, beside = FALSE,
        horiz = FALSE, density = NULL, angle = 45,  col = NULL,
       border = par("fg"), xlab ="COUNTY", ylab = "COUNT")
```

the result is:



The highest number of vaccine provider is Jefferson county, 28 providers.

The lowest number of vaccine provider is Coosa and Wilcox county, 1 providers.

```
> summary(provider_count$count)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  1.000   3.500   5.000   6.075   7.000  28.000
```

In this summary, in 67 counties, median number of providers is 5, the average number of providers is 6, the 1st quantile number of providers is 3.5, the 3rd quantile number of providers is 7.

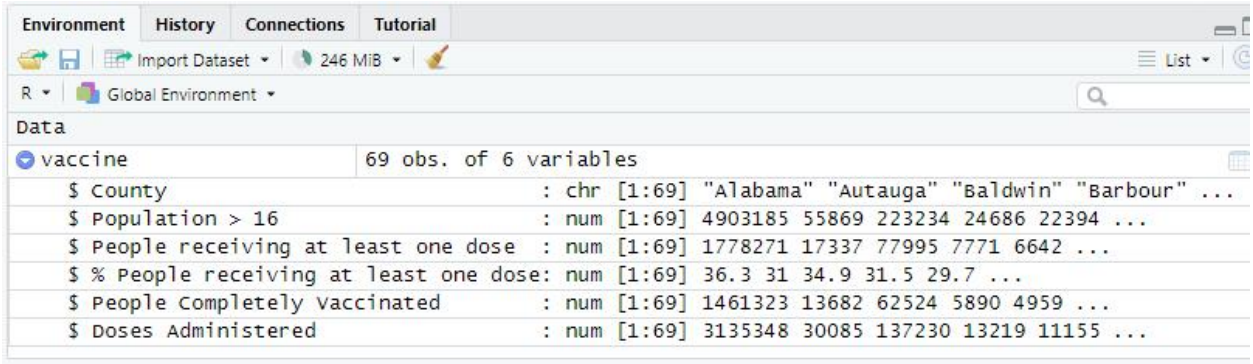
Question3：

Code explanation:

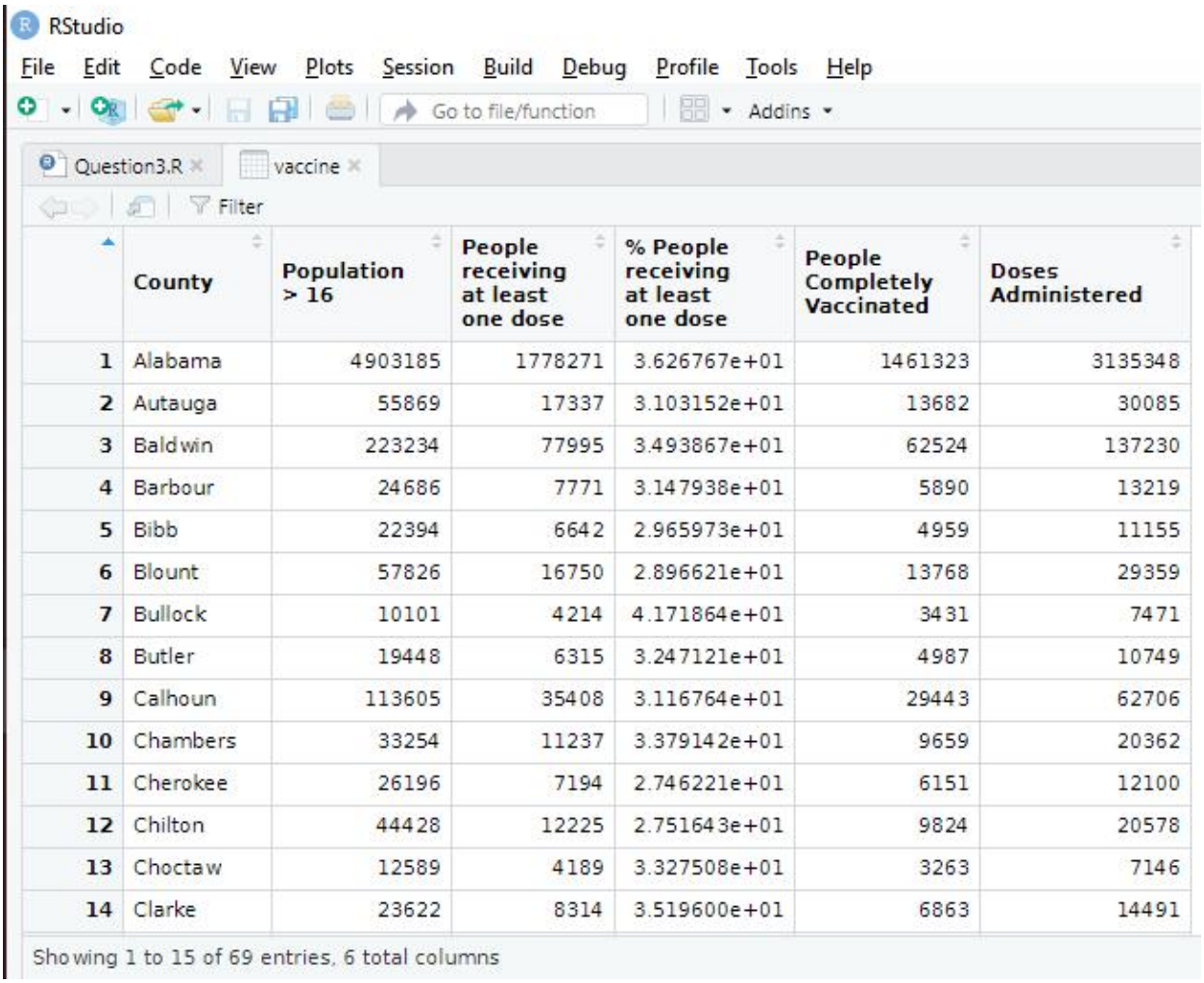
First I install the package “readxl”, then recall it using library(readxl)

Then I want to read the excel file and assign it to the new vector named “vaccine”: vaccine<-read_xlsx("C:/Users/USER/Downloads/COVID-19 County Vaccine Data.xlsx").

After you run this code, at the environment area of R studio, you can see this data named”vaccine”, 69 observations of 6 variables:



If you click the “vaccine”button, you can see the table has already imported into R studio:



	County	Population > 16	People receiving at least one dose	% People receiving at least one dose	People Completely Vaccinated	Doses Administered
1	Alabama	4903185	1778271	3.626767e+01	1461323	3135348
2	Autauga	55869	17337	3.103152e+01	13682	30085
3	Baldwin	223234	77995	3.493867e+01	62524	137230
4	Barbour	24686	7771	3.147938e+01	5890	13219
5	Bibb	22394	6642	2.965973e+01	4959	11155
6	Blount	57826	16750	2.896621e+01	13768	29359
7	Bullock	10101	4214	4.171864e+01	3431	7471
8	Butler	19448	6315	3.247121e+01	4987	10749
9	Calhoun	113605	35408	3.116764e+01	29443	62706
10	Chambers	33254	11237	3.379142e+01	9659	20362
11	Cherokee	26196	7194	2.746221e+01	6151	12100
12	Chilton	44428	12225	2.751643e+01	9824	20578
13	Choctaw	12589	4189	3.327508e+01	3263	7146
14	Clarke	23622	8314	3.519600e+01	6863	14491

Showing 1 to 15 of 69 entries, 6 total columns

Then I want to see how many N/A values in this dataset, I use code: table(is.na(vaccine)) to see how many N/A values in vaccine dataset. After you run this code, you can see the result:

```
FALSE
  414
```

False means no N/A values, and 414 False means there are 414 values in the dataset. Since no true, we don't have missing values.

Since there is a -1 in the "poplation>16"group, I think it is not right, cause our data should greater than 16, so I name the numbers less than 16 as N/A, using code: vaccine$`Population > 16`[which(vaccine$`Population > 16`<=16)]=NA;

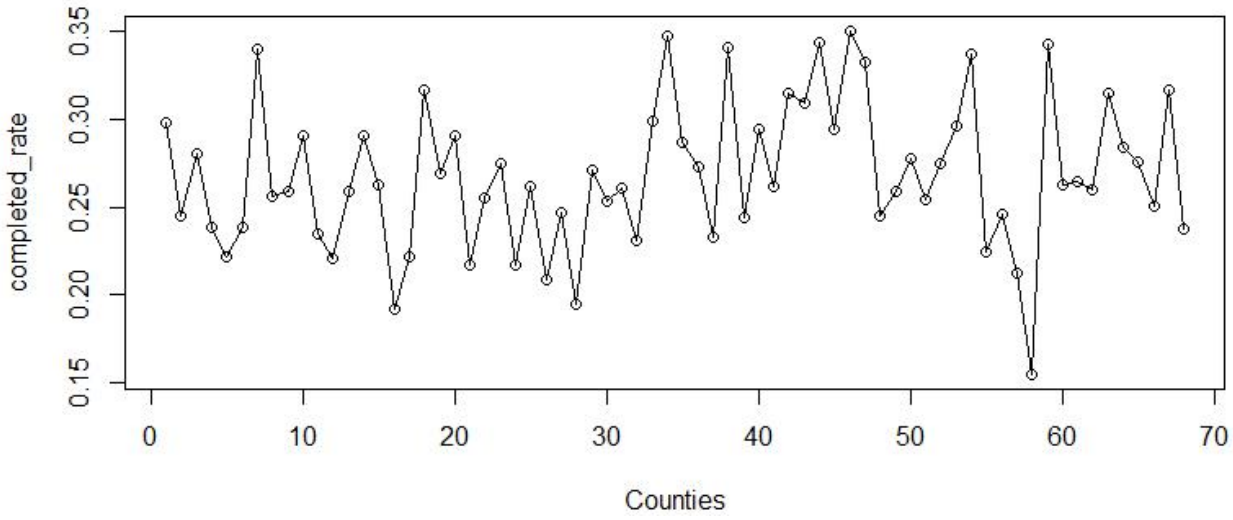
And then I want to ask R studio, is that value N/A? so I use code: is.na(vaccine), then I found that value who less than 16 is N/A.

[61,]	FALSE	FALSE	FALSE	FALSE
[62,]	FALSE	FALSE	FALSE	FALSE
[63,]	FALSE	FALSE	FALSE	FALSE
[64,]	FALSE	FALSE	FALSE	FALSE
[65,]	FALSE	TRUE	FALSE	FALSE
[66,]	FALSE	FALSE	FALSE	FALSE
[67,]	FALSE	FALSE	FALSE	FALSE
[68,]	FALSE	FALSE	FALSE	FALSE
[69,]	FALSE	FALSE	FALSE	FALSE

Now I want to delete/omit that value, cause it is a N/A, I use code: Vaccine<-na.omit(vaccine). Note that now it is capital V in vaccine.

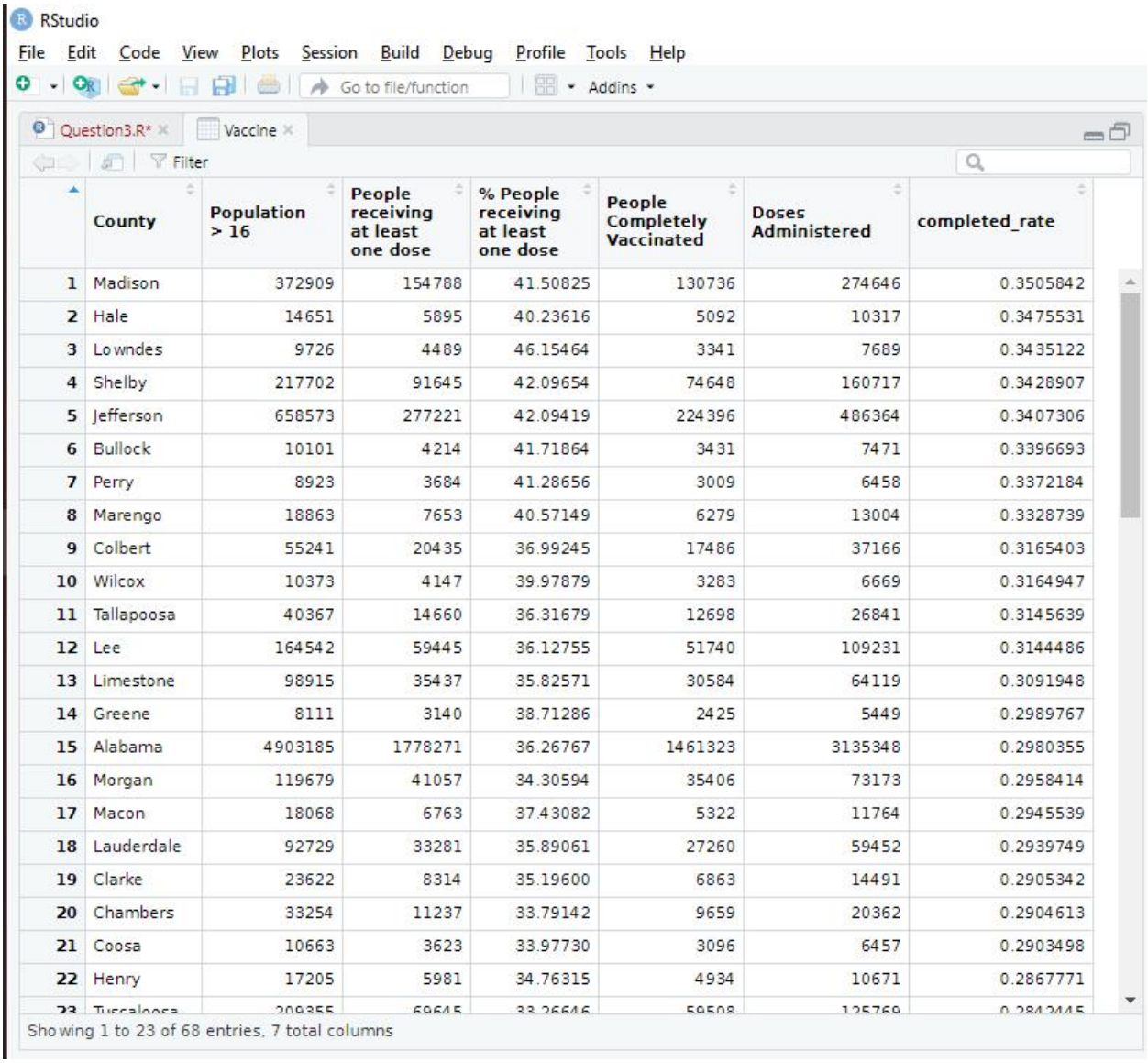
Then I want to calculate completed rate, it equals to people completely vaccinated divided by population>16, I use code: Vaccine$completed_rate<-as.numeric(Vaccine$`People Completely Vaccinated`)/as.numeric(Vaccine$`Population > 16`)

Then I plot the completed rate line chart, I assign county as independent value and completed rate as dependent value, the code is: plot(Vaccine$completed_rate,type="o",xlab = "Counties",ylab = "completed_rate"). The line chart is bottom:



Our counties are in the range of 1-69, each number represents one county, and our completed rate are in the range of 0.1544 to 0.3505.

Then I want to sort the completed rate by the decreasing order, I use code: Vaccine<-Vaccine[order(Vaccine$completed_rate, decreasing= T),] Now, when you click the right side, environment group, second Vaccine button (Capital V), you will get the table automatically, and also get the completed rate in the decreasing order, the table is at bottom:


RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

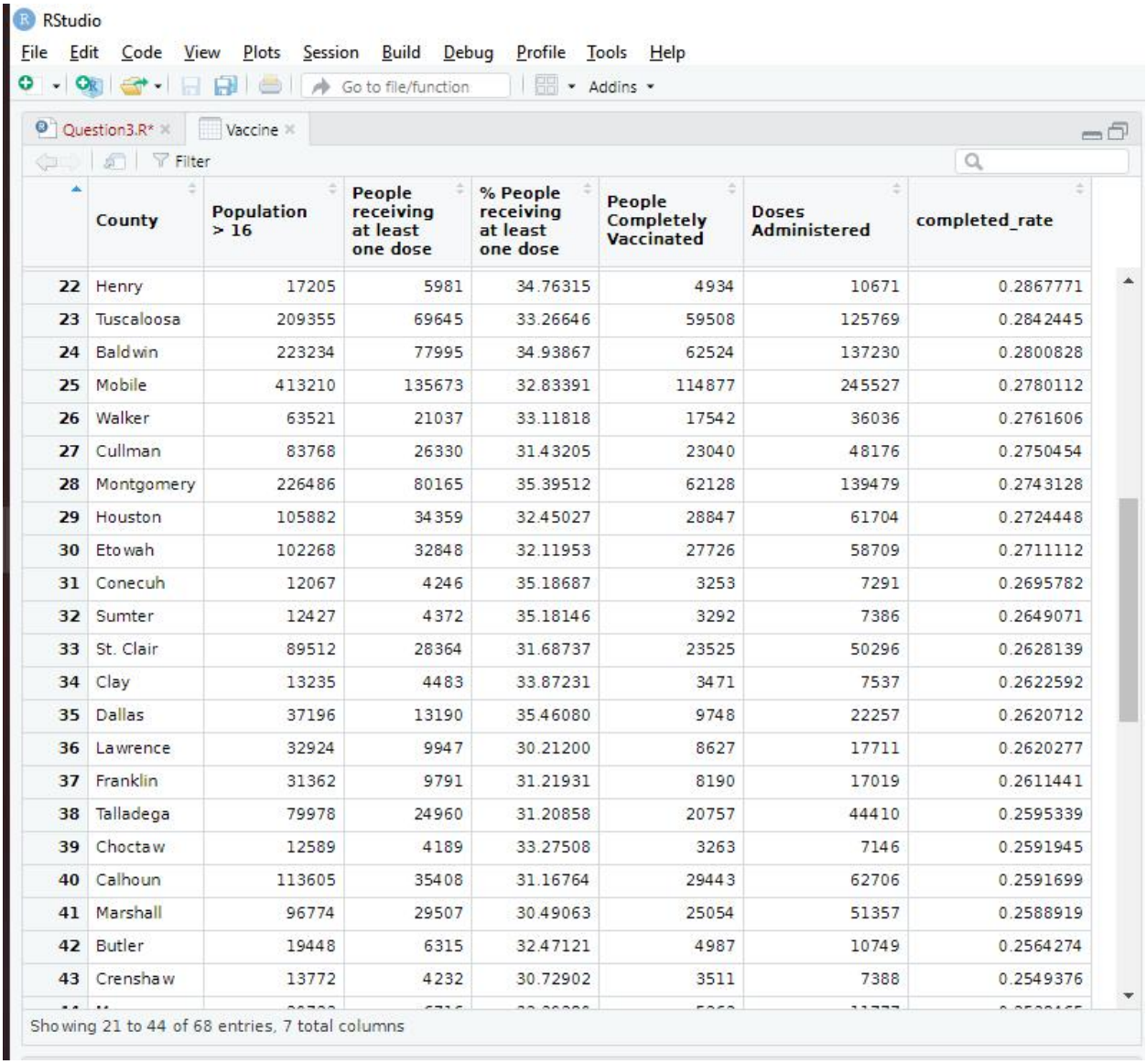
Go to file/function | Addins

Question3.R* | Vaccine

Filter

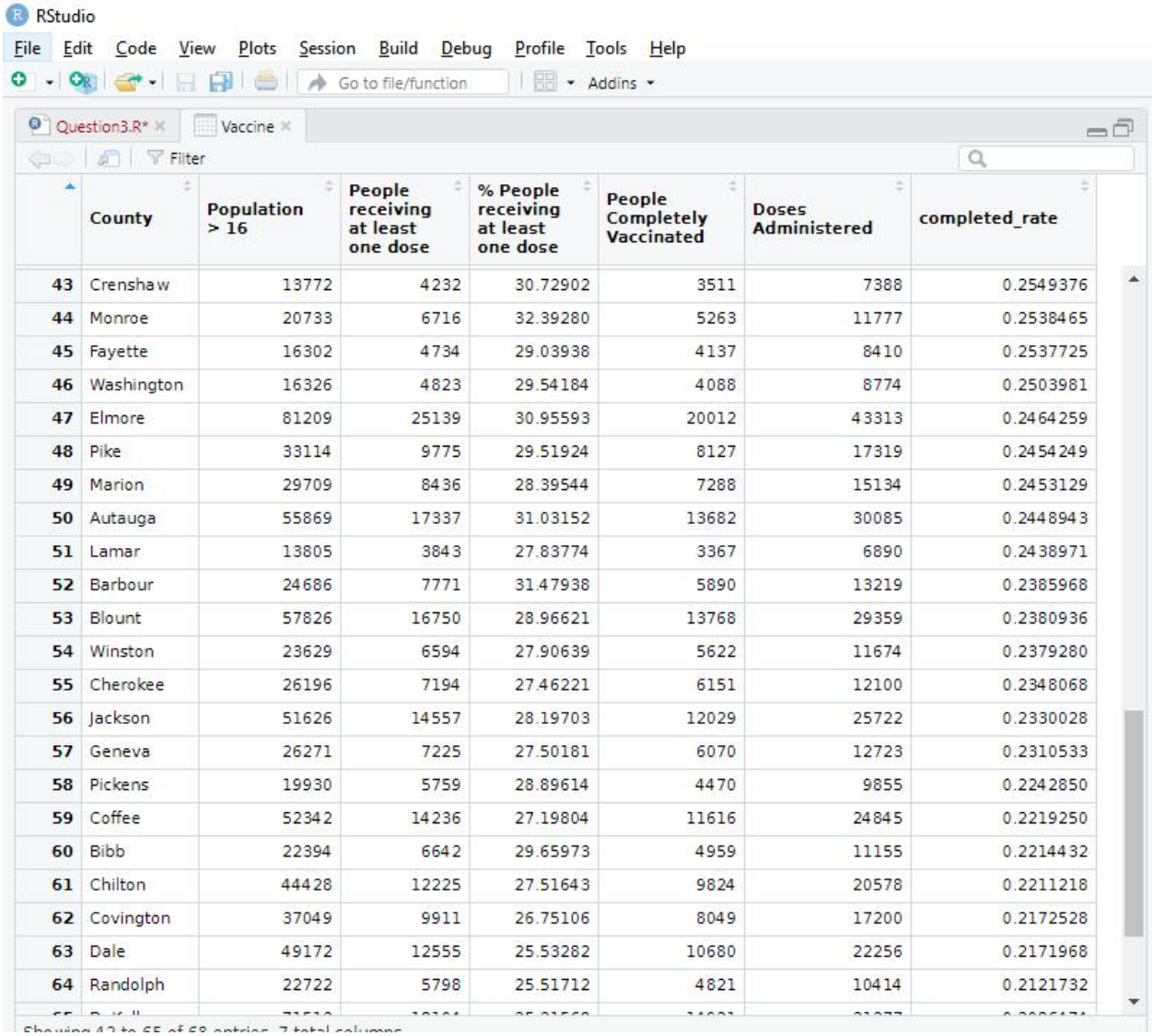
	County	Population > 16	People receiving at least one dose	% People receiving at least one dose	People Completely Vaccinated	Doses Administered	completed_rate
1	Madison	372909	154788	41.50825	130736	274646	0.3505842
2	Hale	14651	5895	40.23616	5092	10317	0.3475531
3	Lowndes	9726	4489	46.15464	3341	7689	0.3435122
4	Shelby	217702	91645	42.09654	74648	160717	0.3428907
5	Jefferson	658573	277221	42.09419	224396	486364	0.3407306
6	Bullock	10101	4214	41.71864	3431	7471	0.3396693
7	Perry	8923	3684	41.28656	3009	6458	0.3372184
8	Marengo	18863	7653	40.57149	6279	13004	0.3328739
9	Colbert	55241	20435	36.99245	17486	37166	0.3165403
10	Wilcox	10373	4147	39.97879	3283	6669	0.3164947
11	Tallapoosa	40367	14660	36.31679	12698	26841	0.3145639
12	Lee	164542	59445	36.12755	51740	109231	0.3144486
13	Limestone	98915	35437	35.82571	30584	64119	0.3091948
14	Greene	8111	3140	38.71286	2425	5449	0.2989767
15	Alabama	4903185	1778271	36.26767	1461323	3135348	0.2980355
16	Morgan	119679	41057	34.30594	35406	73173	0.2958414
17	Macon	18068	6763	37.43082	5322	11764	0.2945539
18	Lauderdale	92729	33281	35.89061	27260	59452	0.2939749
19	Clarke	23622	8314	35.19600	6863	14491	0.2905342
20	Chambers	33254	11237	33.79142	9659	20362	0.2904613
21	Coosa	10663	3623	33.97730	3096	6457	0.2903498
22	Henry	17205	5981	34.76315	4934	10671	0.2867771
23	Tuscaloosa	209355	69645	33.26646	59508	125769	0.2842445

Showing 1 to 23 of 68 entries, 7 total columns

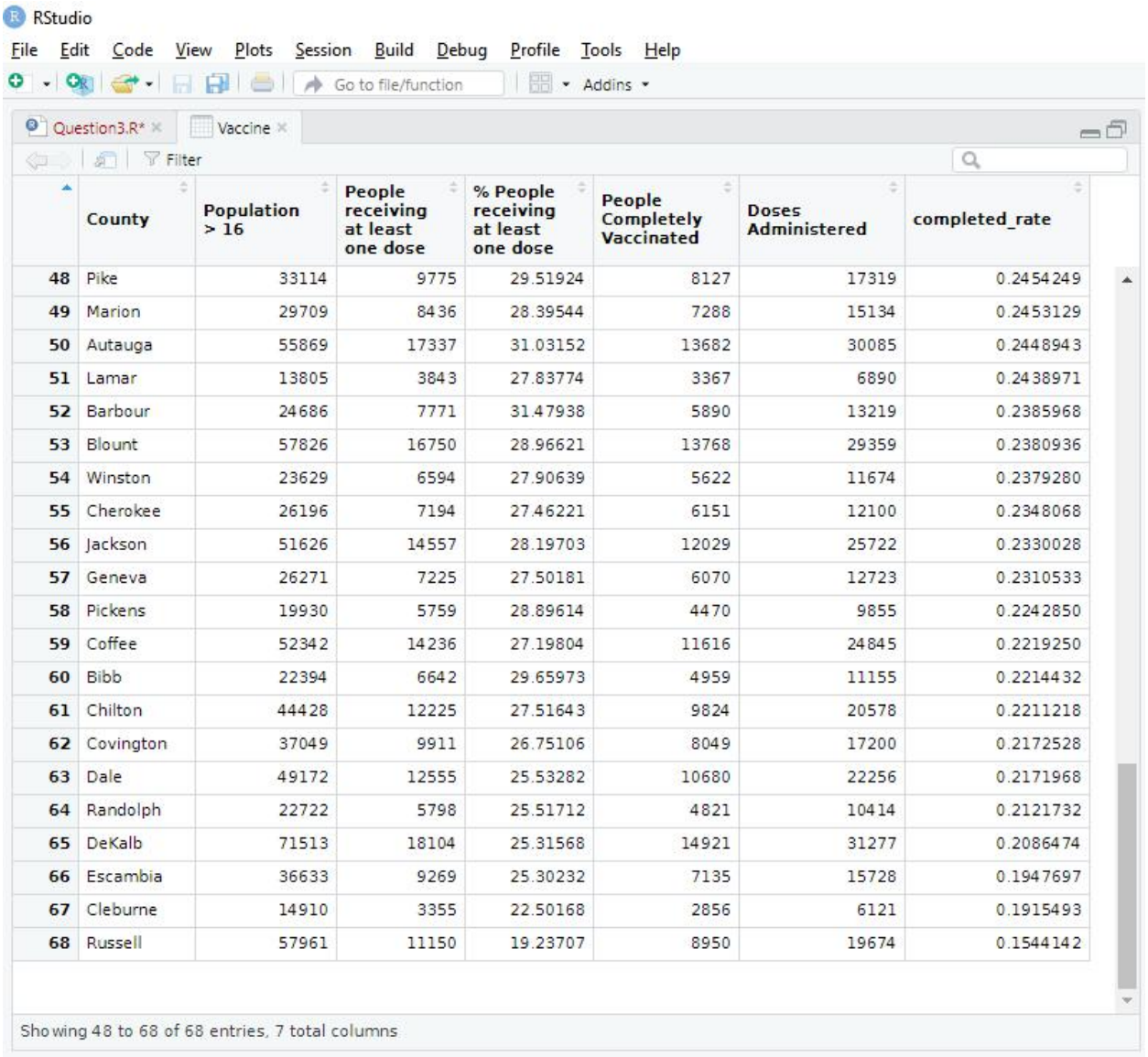


	County	Population > 16	People receiving at least one dose	% People receiving at least one dose	People Completely Vaccinated	Doses Administered	completed_rate
22	Henry	17205	5981	34.76315	4934	10671	0.2867771
23	Tuscaloosa	209355	69645	33.26646	59508	125769	0.2842445
24	Baldwin	223234	77995	34.93867	62524	137230	0.2800828
25	Mobile	413210	135673	32.83391	114877	245527	0.2780112
26	Walker	63521	21037	33.11818	17542	36036	0.2761606
27	Cullman	83768	26330	31.43205	23040	48176	0.2750454
28	Montgomery	226486	80165	35.39512	62128	139479	0.2743128
29	Houston	105882	34359	32.45027	28847	61704	0.2724448
30	Etowah	102268	32848	32.11953	27726	58709	0.2711112
31	Conecuh	12067	4246	35.18687	3253	7291	0.2695782
32	Sumter	12427	4372	35.18146	3292	7386	0.2649071
33	St. Clair	89512	28364	31.68737	23525	50296	0.2628139
34	Clay	13235	4483	33.87231	3471	7537	0.2622592
35	Dallas	37196	13190	35.46080	9748	22257	0.2620712
36	Lawrence	32924	9947	30.21200	8627	17711	0.2620277
37	Franklin	31362	9791	31.21931	8190	17019	0.2611441
38	Talladega	79978	24960	31.20858	20757	44410	0.2595339
39	Choctaw	12589	4189	33.27508	3263	7146	0.2591945
40	Calhoun	113605	35408	31.16764	29443	62706	0.2591699
41	Marshall	96774	29507	30.49063	25054	51357	0.2588919
42	Butler	19448	6315	32.47121	4987	10749	0.2564274
43	Crenshaw	13772	4232	30.72902	3511	7388	0.2549376

Showing 21 to 44 of 68 entries, 7 total columns



	County	Population > 16	People receiving at least one dose	% People receiving at least one dose	People Completely Vaccinated	Doses Administered	completed_rate
43	Crenshaw	13772	4232	30.72902	3511	7388	0.2549376
44	Monroe	20733	6716	32.39280	5263	11777	0.2538465
45	Fayette	16302	4734	29.03938	4137	8410	0.2537725
46	Washington	16326	4823	29.54184	4088	8774	0.2503981
47	Elmore	81209	25139	30.95593	20012	43313	0.2464259
48	Pike	33114	9775	29.51924	8127	17319	0.2454249
49	Marion	29709	8436	28.39544	7288	15134	0.2453129
50	Autauga	55869	17337	31.03152	13682	30085	0.2448943
51	Lamar	13805	3843	27.83774	3367	6890	0.2438971
52	Barbour	24686	7771	31.47938	5890	13219	0.2385968
53	Blount	57826	16750	28.96621	13768	29359	0.2380936
54	Winston	23629	6594	27.90639	5622	11674	0.2379280
55	Cherokee	26196	7194	27.46221	6151	12100	0.2348068
56	Jackson	51626	14557	28.19703	12029	25722	0.2330028
57	Geneva	26271	7225	27.50181	6070	12723	0.2310533
58	Pickens	19930	5759	28.89614	4470	9855	0.2242850
59	Coffee	52342	14236	27.19804	11616	24845	0.2219250
60	Bibb	22394	6642	29.65973	4959	11155	0.2214432
61	Chilton	44428	12225	27.51643	9824	20578	0.2211218
62	Covington	37049	9911	26.75106	8049	17200	0.2172528
63	Dale	49172	12555	25.53282	10680	22256	0.2171968
64	Randolph	22722	5798	25.51712	4821	10414	0.2121732



	County	Population > 16	People receiving at least one dose	% People receiving at least one dose	People Completely Vaccinated	Doses Administered	completed_rate
48	Pike	33114	9775	29.51924	8127	17319	0.2454249
49	Marion	29709	8436	28.39544	7288	15134	0.2453129
50	Autauga	55869	17337	31.03152	13682	30085	0.2448943
51	Lamar	13805	3843	27.83774	3367	6890	0.2438971
52	Barbour	24686	7771	31.47938	5890	13219	0.2385968
53	Blount	57826	16750	28.96621	13768	29359	0.2380936
54	Winston	23629	6594	27.90639	5622	11674	0.2379280
55	Cherokee	26196	7194	27.46221	6151	12100	0.2348068
56	Jackson	51626	14557	28.19703	12029	25722	0.2330028
57	Geneva	26271	7225	27.50181	6070	12723	0.2310533
58	Pickens	19930	5759	28.89614	4470	9855	0.2242850
59	Coffee	52342	14236	27.19804	11616	24845	0.2219250
60	Bibb	22394	6642	29.65973	4959	11155	0.2214432
61	Chilton	44428	12225	27.51643	9824	20578	0.2211218
62	Covington	37049	9911	26.75106	8049	17200	0.2172528
63	Dale	49172	12555	25.53282	10680	22256	0.2171968
64	Randolph	22722	5798	25.51712	4821	10414	0.2121732
65	DeKalb	71513	18104	25.31568	14921	31277	0.2086474
66	Escambia	36633	9269	25.30232	7135	15728	0.1947697
67	Cleburne	14910	3355	22.50168	2856	6121	0.1915493
68	Russell	57961	11150	19.23707	8950	19674	0.1544142

Showing 48 to 68 of 68 entries, 7 total columns

As you can see in the rightmost column of this table, the completed rate is ordered by decreasing order.

Then I want to do the summary statistics, summary(Vaccine), the result is as follows:

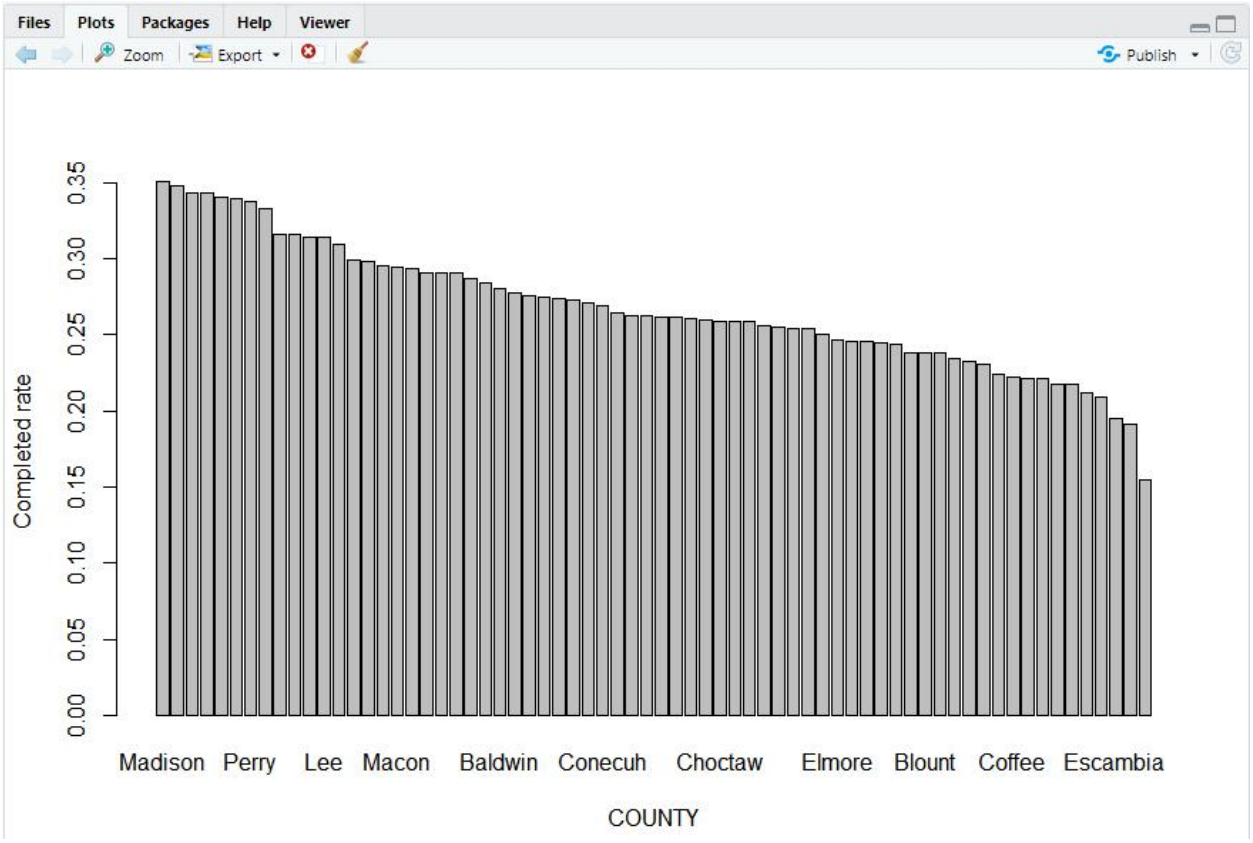
```
> summary(Vaccine)
    County          Population > 16   People receiving at least one dose
 Length:68          Min.   :   8111   Min.   :   3140
 Class :character   1st Qu.:  17852   1st Qu.:   5871
 Mode  :character   Median :  34944   Median :   9929
                    Mean   : 144211   Mean   :  51211
                    3rd Qu.:  85204   3rd Qu.:  26839
                    Max.   :4903185   Max.   :1778271
 % People receiving at least one dose People Completely Vaccinated Doses Administered
 Min.   :19.24                        Min.   :   2425              Min.   :   5449
 1st Qu.:29.02                        1st Qu.:   4906              1st Qu.:  10390
 Median :32.42                        Median :   8408              Median :  17515
 Mean   :32.69                        Mean   :  42207              Mean   :  90432
 3rd Qu.:35.55                        3rd Qu.:  23161              3rd Qu.:  48706
 Max.   :46.15                        Max.   :1461323              Max.   :3135348
 completed_rate
 Min.   :0.1544
 1st Qu.:0.2426
 Median :0.2622
 Mean   :0.2681
 3rd Qu.:0.2941
 Max.   :0.3506
```

	Population>16	People receiving at least one dose	People Completely Vaccinted	Completled rate
min	8111	19.24	2425	0.1544
1st Qu	17852	29.02	4906	0.2426
Median	34944	32.42	8404	0.2622
Mean	144211	32.69	42207	0.2681
3rd Qu	85204	26839	23161	0.2941
Max	4903185	46.15	1461323	0.3506

Then I want to see the bar chart of completed rate by county, code:

```
barplot(Vaccine$completed_rate, names.arg =Vaccine$County , beside = FALSE,
    horiz = FALSE, density = NULL, angle = 45,  col = NULL,
    border = par("fg"), xlab ="COUNTY", ylab = "Completed rate")
```

result is:



Due to the fact that we have 68 counties, cannot fit the bar chart, so it only gives us several names at equal intervals.

Madison county has the highest completed rate, which is 35.05%,

Russell county has the lowest completed rate, which is 15.44%.

Mean completed rate is 26.81%, Pike county is little under the mean, has 24.54%.

Clay county is at the median level of completed rate, which is 26.22%.