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.

Graphical user interface, text, application

Description automatically generated

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

Graphical user interface, text, application, email

Description automatically generated

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:



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

Application

Description automatically generated with medium confidence

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

Graphical user interface, table

Description automatically generated with medium confidence

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:

Graphical user interface

Description automatically generated with medium confidence

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.

Graphical user interface, text, application

Description automatically generatedIn 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;

Graphical user interface, application

Description automatically generated

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

Graphical user interface, text, application

Description automatically generatedUntil 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,

Graphical user interface, text, application, email

Description automatically generated

Then you click the provider\_count button:

Table

Description automatically generated with medium confidence

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:

Table

Description automatically generated with medium confidence

Table

Description automatically generated with medium confidence

Table

Description automatically generated with medium confidence

Graphical user interface, table

Description automatically generated with medium confidence

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:

Chart

Description automatically generated

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

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



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.