Example Manuscript Template for a Data Analysis Project

Andreas Handel

2021-10-27

#Loading required packages  
library(dplyr) #for data processing

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(here) #to set paths

## here() starts at C:/Users/Priyanka/Desktop/new/Priyanka\_G-MADA-project

library(tidyverse) #all required data manipulation packages

## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5 v purrr 0.3.4  
## v tibble 3.1.3 v stringr 1.4.0  
## v tidyr 1.1.4 v forcats 0.5.1  
## v readr 2.0.1

## -- Conflicts ------------------------------------------ tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

# 1 Summary/Abstract

*Write a summary of your project.*

# 2 KEY QUESTIONS

What is already known about this subject?

Residents of rural communities are at increased risk for severe COVID-19–associated morbidity and mortality. In September 2020, COVID-19 incidence (cases per 100,000 population) in rural counties surpassed that in urban counties.

What will this study add or answer?

The present analysis will attempt to estimate the percent of the population in each county that may be vaccine hesitant

Relationship between ethnicity and vaccine hesitancy

How might this impact clinical practice?

Identify and address barriers to COVID-19 vaccination in rural areas.

## 2.1 General Background Information

Coronavirus disease (COVID-19) is an infectious disease caused by a newly discovered coronavirus. Coronavirus disease 2019 (COVID-19) was declared a pandemic in March 2020. COVID-19 vaccine is the most sustainable option to manage the current pandemic. However, vaccine hesitancy by even a small subset of the population can undermine the success of this strategy.

## 2.2 Description of data and data source

*Describe what the data is, what it contains, where it is from, etc. Eventually this might be part of a methods section.*

The dataset used for this analysis is publicly available on CDC website. <https://data.cdc.gov/Vaccinations/Vaccine-Hesitancy-for-COVID-19-County-and-local-es/q9mh-h2tw>

The hesitancy rates at the county level using the Public Use Microdata Sample (PUMS). The data was collected by utilizing survey question, “Once a vaccine to prevent COVID-19 is available to you, would you…get a vaccine?” which provides the following options:

1. “definitely get a vaccine”;
2. “probably get a vaccine”;
3. “unsure”; 4) “probably not get a vaccine”;
4. “definitely not get a vaccine.”

We use three definitions to capture the strength of hesitancy to receive a vaccine.

Strongly hesitant: includes only survey responses indicating that they would “definitely not” receive a COVID-19 vaccine when available.

Hesitant: includes survey responses indicating that they would “probably not” or “definitely not” receive a COVID-19 vaccine when available.

Hesitant or unsure: includes survey responses indicating that they would “probably not” or “unsure” or “definitely not” receive a COVID-19 vaccine when available

## 2.3 Questions/Hypotheses to be addressed

*State the research questions you plan to answer with this analysis.*

The present analysis will attempt to estimate the percent of the population in each county that may be vaccine hesitant

Relationship between ethnicity and vaccine hesitancy

# 3 Future analysis

I plan to present a summary of county-wise vaccine estimate based on race and ethnicity. I will plot some graphs (box-plot, scatter plot). I am also planning to do some regression analysis too

# 4 Methods and Results

*In most research papers, results and methods are separate. You can combine them here if you find it easier. You are also welcome to structure things such that those are separate sections.*

## 4.1 Data aquisition

*As applicable, explain where and how you got the data. If you directly import the data from an online source, you can combine this section with the next.*

## 4.2 Data import and cleaning

*Write code that reads in the file and cleans it so it’s ready for analysis. Since this will be fairly long code for most datasets, it might be a good idea to have it in one or several R scripts. If that is the case, explain here briefly what kind of cleaning/processing you do, and provide more details and well documented code somewhere (e.g. as supplement in a paper). All materials, including files that contain code, should be commented well so everyone can follow along.*

library(readxl) #for loading Excel files  
library(dplyr) #for data processing  
library(here) #to set paths  
library(tidyverse)

#path to data

Covid\_data\_location <- here::here("data","raw\_data","Vaccine\_Hesitancy\_for\_COVID-19.xlsx")

#load data.

rawdata <- readxl::read\_excel(Covid\_data\_location)

#take a look at the data

dplyr::glimpse(rawdata)

## Rows: 3,142  
## Columns: 21  
## $ `FIPS Code` <dbl> 112~  
## $ `County Name` <chr> "Ta~  
## $ State <chr> "AL~  
## $ `Estimated hesitant` <dbl> 0.1~  
## $ `Estimated hesitant or unsure` <dbl> 0.2~  
## $ `Estimated strongly hesitant` <dbl> 0.1~  
## $ `Social Vulnerability Index (SVI)` <dbl> 0.8~  
## $ `SVI Category` <chr> "Ve~  
## $ `CVAC level of concern for vaccination rollout` <dbl> 0.6~  
## $ `CVAC Level Of Concern` <chr> "Hi~  
## $ `Percent adults fully vaccinated against COVID-19 (as of 6/10/21)` <dbl> 0.3~  
## $ `Percent Hispanic` <dbl> 0.0~  
## $ `Percent non-Hispanic American Indian/Alaska Native` <dbl> 0.0~  
## $ `Percent non-Hispanic Asian` <dbl> 0.0~  
## $ `Percent non-Hispanic Black` <dbl> 0.2~  
## $ `Percent non-Hispanic Native Hawaiian/Pacific Islander` <dbl> 0.0~  
## $ `Percent non-Hispanic White` <dbl> 0.6~  
## $ `Geographical Point` <chr> "PO~  
## $ `State Code` <chr> "AL~  
## $ `County Boundary` <chr> "MU~  
## $ `State Boundary` <chr> "MU~

str(rawdata)

## tibble [3,142 x 21] (S3: tbl\_df/tbl/data.frame)  
## $ FIPS Code : num [1:3142] 1123 1121 1131 1129 1133 ...  
## $ County Name : chr [1:3142] "Tallapoosa County, Alabama" "Talladega County, Alabama" "Wilcox County, Alabama" "Washington County, Alabama" ...  
## $ State : chr [1:3142] "ALABAMA" "ALABAMA" "ALABAMA" "ALABAMA" ...  
## $ Estimated hesitant : num [1:3142] 0.181 0.178 0.173 0.173 0.18 ...  
## $ Estimated hesitant or unsure : num [1:3142] 0.24 0.235 0.236 0.236 0.231 ...  
## $ Estimated strongly hesitant : num [1:3142] 0.138 0.137 0.134 0.134 0.138 ...  
## $ Social Vulnerability Index (SVI) : num [1:3142] 0.89 0.87 0.93 0.73 0.7 0.75 0.58 0.49 0.63 0.97 ...  
## $ SVI Category : chr [1:3142] "Very High Vulnerability" "Very High Vulnerability" "Very High Vulnerability" "High Vulnerability" ...  
## $ CVAC level of concern for vaccination rollout : num [1:3142] 0.64 0.84 0.94 0.82 0.8 0.68 0.87 0.77 0.58 0.89 ...  
## $ CVAC Level Of Concern : chr [1:3142] "High Concern" "Very High Concern" "Very High Concern" "Very High Concern" ...  
## $ Percent adults fully vaccinated against COVID-19 (as of 6/10/21): num [1:3142] 0.305 0.265 0.394 0.308 0.163 0.357 0.672 0.475 0.328 0.689 ...  
## $ Percent Hispanic : num [1:3142] 0.0242 0.0229 0.0053 0.0146 0.0315 ...  
## $ Percent non-Hispanic American Indian/Alaska Native : num [1:3142] 0.0022 0.0043 0.0009 0.0731 0.0034 ...  
## $ Percent non-Hispanic Asian : num [1:3142] 0.0036 0.0061 0.0003 0.0025 0.0016 ...  
## $ Percent non-Hispanic Black : num [1:3142] 0.2697 0.3237 0.6938 0.2354 0.0073 ...  
## $ Percent non-Hispanic Native Hawaiian/Pacific Islander : num [1:3142] 0 0.0003 0 0 0.0005 0 0.01 0.0265 0.0004 0.0022 ...  
## $ Percent non-Hispanic White : num [1:3142] 0.689 0.626 0.268 0.649 0.937 ...  
## $ Geographical Point : chr [1:3142] "POINT (-86.844516 32.756889)" "POINT (-86.844516 32.756889)" "POINT (-86.844516 32.756889)" "POINT (-86.844516 32.756889)" ...  
## $ State Code : chr [1:3142] "AL" "AL" "AL" "AL" ...  
## $ County Boundary : chr [1:3142] "MULTIPOLYGON (((-85.841259 33.104456, -85.84098999999999 33.104456, -85.840717 33.104455, -85.837858 33.104441,"| \_\_truncated\_\_ "MULTIPOLYGON (((-86.303069 33.46316, -86.30306999999999 33.463223, -86.303116 33.463974, -86.30333999999999 33."| \_\_truncated\_\_ "MULTIPOLYGON (((-87.52534299999999 32.132773, -87.521946 32.132816, -87.52191599999999 32.150816999999996, -87."| \_\_truncated\_\_ "MULTIPOLYGON (((-88.45317899999999 31.505388, -88.453299 31.509179999999997, -88.45337099999999 31.511405999999"| \_\_truncated\_\_ ...  
## $ State Boundary : chr [1:3142] "MULTIPOLYGON (((-88.139988 34.581703, -88.13526 34.616806, -88.11840699999999 34.724292, -88.116418 34.746303, "| \_\_truncated\_\_ "MULTIPOLYGON (((-88.139988 34.581703, -88.13526 34.616806, -88.11840699999999 34.724292, -88.116418 34.746303, "| \_\_truncated\_\_ "MULTIPOLYGON (((-88.139988 34.581703, -88.13526 34.616806, -88.11840699999999 34.724292, -88.116418 34.746303, "| \_\_truncated\_\_ "MULTIPOLYGON (((-88.139988 34.581703, -88.13526 34.616806, -88.11840699999999 34.724292, -88.116418 34.746303, "| \_\_truncated\_\_ ...

summary(rawdata)

## FIPS Code County Name State Estimated hesitant  
## Min. : 1001 Length:3142 Length:3142 Min. :0.0269   
## 1st Qu.:18178 Class :character Class :character 1st Qu.:0.0983   
## Median :29176 Mode :character Mode :character Median :0.1318   
## Mean :30384 Mean :0.1326   
## 3rd Qu.:45081 3rd Qu.:0.1617   
## Max. :56045 Max. :0.2670   
##   
## Estimated hesitant or unsure Estimated strongly hesitant  
## Min. :0.0499 Min. :0.01860   
## 1st Qu.:0.1485 1st Qu.:0.06232   
## Median :0.1901 Median :0.08490   
## Mean :0.1914 Mean :0.08665   
## 3rd Qu.:0.2288 3rd Qu.:0.10447   
## Max. :0.3233 Max. :0.18240   
##   
## Social Vulnerability Index (SVI) SVI Category   
## Min. :0.00 Length:3142   
## 1st Qu.:0.25 Class :character   
## Median :0.50 Mode :character   
## Mean :0.50   
## 3rd Qu.:0.75   
## Max. :1.00   
## NA's :1   
## CVAC level of concern for vaccination rollout CVAC Level Of Concern  
## Min. :0.00 Length:3142   
## 1st Qu.:0.25 Class :character   
## Median :0.50 Mode :character   
## Mean :0.50   
## 3rd Qu.:0.75   
## Max. :1.00   
##   
## Percent adults fully vaccinated against COVID-19 (as of 6/10/21)  
## Min. :0.0010   
## 1st Qu.:0.3180   
## Median :0.4000   
## Mean :0.3994   
## 3rd Qu.:0.4940   
## Max. :0.9990   
## NA's :278   
## Percent Hispanic Percent non-Hispanic American Indian/Alaska Native  
## Min. :0.00000 Min. :0.00000   
## 1st Qu.:0.02220 1st Qu.:0.00120   
## Median :0.04230 Median :0.00280   
## Mean :0.09418 Mean :0.01846   
## 3rd Qu.:0.09680 3rd Qu.:0.00690   
## Max. :0.99170 Max. :0.91900   
##   
## Percent non-Hispanic Asian Percent non-Hispanic Black  
## Min. :0.00000 Min. :0.00000   
## 1st Qu.:0.00280 1st Qu.:0.00650   
## Median :0.00610 Median :0.02190   
## Mean :0.01362 Mean :0.08926   
## 3rd Qu.:0.01280 3rd Qu.:0.09840   
## Max. :0.41730 Max. :0.87230   
##   
## Percent non-Hispanic Native Hawaiian/Pacific Islander  
## Min. :0.0000000   
## 1st Qu.:0.0000000   
## Median :0.0001000   
## Mean :0.0009189   
## 3rd Qu.:0.0006000   
## Max. :0.2727000   
##   
## Percent non-Hispanic White Geographical Point State Code   
## Min. :0.0069 Length:3142 Length:3142   
## 1st Qu.:0.6444 Class :character Class :character   
## Median :0.8374 Mode :character Mode :character   
## Mean :0.7625   
## 3rd Qu.:0.9249   
## Max. :1.0000   
##   
## County Boundary State Boundary   
## Length:3142 Length:3142   
## Class :character Class :character   
## Mode :character Mode :character   
##   
##   
##   
##

#Data cleaning

##renaming Columns

rawdata <- rawdata %>%  
 rename(Hispanic = `Percent Hispanic`)  
  
rawdata <- rawdata %>%  
 rename(Asian = `Percent non-Hispanic Asian`)  
  
 rawdata <- rawdata %>%  
 rename(Black =`Percent non-Hispanic Black`)  
  
 rawdata <- rawdata %>%  
 rename(White ='Percent non-Hispanic White')  
   
##Recoding Data  
   
   
 rawdata <- rawdata %>%  
 mutate(`SVI Category` = case\_when(`SVI Category`=="Very Low Vulnerability" ~ 1,  
 `SVI Category`=="Low Vulnerability" ~ 2,  
 `SVI Category`=="Moderate Vulnerability" ~ 3,  
 `SVI Category`=="High Vulnerability" ~ 4,  
 `SVI Category`=="Very High Vulnerability" ~ 5))  
   
   
   
 rawdata <- rawdata %>%  
 mutate( `CVAC Level Of Concern` = case\_when( `CVAC Level Of Concern`=="Very Low Concern" ~ 1,  
 `CVAC Level Of Concern`=="Low Concern" ~ 2,  
 `CVAC Level Of Concern`=="Moderate Concern" ~ 3,  
 `CVAC Level Of Concern`=="High Concern" ~ 4,  
 `CVAC Level Of Concern`=="Very High Concern" ~ 5))

##Creating new variables for Ease

Northeast\_state\_data <- rawdata %>% dplyr::filter(State %in% c("MAINE", "NEW HAMPSHIRE", "VERMONT","MASSACHUSETTS","RHODE ISLAND","CONNECTICUT","NEW YORK","NEW JERSEY","PENNSYLVANIA" ))  
   
## Cross\_Check  
Northeast\_state\_data

## # A tibble: 217 x 21  
## `FIPS Code` `County Name` State `Estimated hesit~ `Estimated hesitan~  
## <dbl> <chr> <chr> <dbl> <dbl>  
## 1 23027 Waldo County, Mai~ MAINE 0.0948 0.128   
## 2 25005 Bristol County, M~ MASSACH~ 0.0572 0.0853  
## 3 25025 Suffolk County, M~ MASSACH~ 0.0442 0.0693  
## 4 25027 Worcester County,~ MASSACH~ 0.049 0.0729  
## 5 9005 Litchfield County~ CONNECT~ 0.0558 0.0775  
## 6 9003 Hartford County, ~ CONNECT~ 0.0531 0.078   
## 7 9011 New London County~ CONNECT~ 0.0622 0.0904  
## 8 9001 Fairfield County,~ CONNECT~ 0.0499 0.0717  
## 9 9013 Tolland County, C~ CONNECT~ 0.0552 0.0783  
## 10 9009 New Haven County,~ CONNECT~ 0.0579 0.0854  
## # ... with 207 more rows, and 16 more variables:  
## # Estimated strongly hesitant <dbl>, Social Vulnerability Index (SVI) <dbl>,  
## # SVI Category <dbl>, CVAC level of concern for vaccination rollout <dbl>,  
## # CVAC Level Of Concern <dbl>,  
## # Percent adults fully vaccinated against COVID-19 (as of 6/10/21) <dbl>,  
## # Hispanic <dbl>, Percent non-Hispanic American Indian/Alaska Native <dbl>,  
## # Asian <dbl>, Black <dbl>, ...

Midwest\_state\_data <- rawdata %>% dplyr::filter(State %in% c("OHIO", "MICHIGAN", "INDIANA","WISCONSIN","ILLINOIS","MINNESOTA","IOWA","MISSOURI","NORTH DAKOTA","SOUTH DAKOTA","NEBRASKA","KANSAS" ))  
   
## Cross\_Check  
  
Midwest\_state\_data

## # A tibble: 1,055 x 21  
## `FIPS Code` `County Name` State `Estimated hesi~ `Estimated hesi~  
## <dbl> <chr> <chr> <dbl> <dbl>  
## 1 26005 Allegan County, Michigan MICHIGAN 0.107 0.184  
## 2 26001 Alcona County, Michigan MICHIGAN 0.0976 0.172  
## 3 26003 Alger County, Michigan MICHIGAN 0.109 0.190  
## 4 26015 Barry County, Michigan MICHIGAN 0.0962 0.171  
## 5 26013 Baraga County, Michigan MICHIGAN 0.095 0.169  
## 6 26011 Arenac County, Michigan MICHIGAN 0.103 0.181  
## 7 26017 Bay County, Michigan MICHIGAN 0.0932 0.163  
## 8 26035 Clare County, Michigan MICHIGAN 0.110 0.198  
## 9 26027 Cass County, Michigan MICHIGAN 0.099 0.174  
## 10 26037 Clinton County, Michigan MICHIGAN 0.0842 0.146  
## # ... with 1,045 more rows, and 16 more variables:  
## # Estimated strongly hesitant <dbl>, Social Vulnerability Index (SVI) <dbl>,  
## # SVI Category <dbl>, CVAC level of concern for vaccination rollout <dbl>,  
## # CVAC Level Of Concern <dbl>,  
## # Percent adults fully vaccinated against COVID-19 (as of 6/10/21) <dbl>,  
## # Hispanic <dbl>, Percent non-Hispanic American Indian/Alaska Native <dbl>,  
## # Asian <dbl>, Black <dbl>, ...

South\_state\_data <- rawdata %>% dplyr::filter(State %in% c("DELAWARE", "MARYLAND", "VIRGINIA"," WEST VIRGINIA","KENTUCKY","NORTH CAROLINA","SOUTH CAROLINA","TENNESSEE","GEORGIA","FLORIDA","ALABAMA","MISSISSIPPI","Arkansas","LOUISIANA","TEXAS","OKLAHOMA","WASHINGTON"))  
  
  
## Cross\_Check  
South\_state\_data

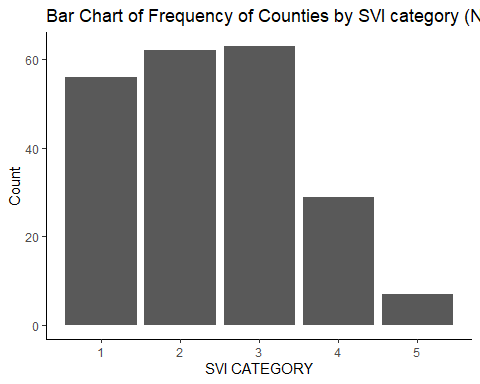
## # A tibble: 1,330 x 21  
## `FIPS Code` `County Name` State `Estimated hesi~ `Estimated hesi~  
## <dbl> <chr> <chr> <dbl> <dbl>  
## 1 1123 Tallapoosa County, Alabama ALABAMA 0.181 0.24   
## 2 1121 Talladega County, Alabama ALABAMA 0.178 0.235  
## 3 1131 Wilcox County, Alabama ALABAMA 0.174 0.236  
## 4 1129 Washington County, Alabama ALABAMA 0.174 0.236  
## 5 1133 Winston County, Alabama ALABAMA 0.180 0.231  
## 6 1127 Walker County, Alabama ALABAMA 0.186 0.240  
## 7 1125 Tuscaloosa County, Alabama ALABAMA 0.16 0.220  
## 8 13165 Jenkins County, Georgia GEORGIA 0.162 0.246  
## 9 13167 Johnson County, Georgia GEORGIA 0.173 0.254  
## 10 13159 Jasper County, Georgia GEORGIA 0.159 0.224  
## # ... with 1,320 more rows, and 16 more variables:  
## # Estimated strongly hesitant <dbl>, Social Vulnerability Index (SVI) <dbl>,  
## # SVI Category <dbl>, CVAC level of concern for vaccination rollout <dbl>,  
## # CVAC Level Of Concern <dbl>,  
## # Percent adults fully vaccinated against COVID-19 (as of 6/10/21) <dbl>,  
## # Hispanic <dbl>, Percent non-Hispanic American Indian/Alaska Native <dbl>,  
## # Asian <dbl>, Black <dbl>, ...

West\_state\_data <- rawdata %>% dplyr::filter(State %in% c("MONTANA", "IDAHO", "WYOMING"," COLORADO","NEW MEXICO","ARIZONA","UTAH","NEVADA","CALIFORNIA","OREGON","ALASKA","HAWAII"))  
  
## Cross\_Check  
  
West\_state\_data

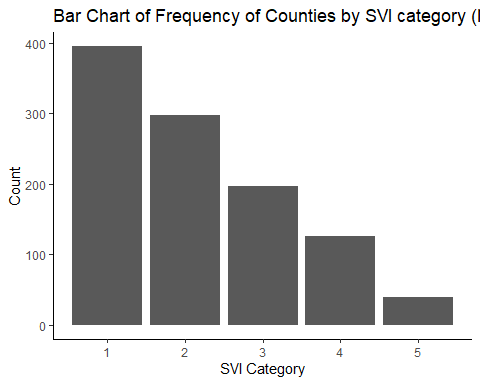
## # A tibble: 345 x 21  
## `FIPS Code` `County Name` State `Estimated hesi~ `Estimated hesi~  
## <dbl> <chr> <chr> <dbl> <dbl>  
## 1 2013 Aleutians East Borough, Alaska ALASKA 0.236 0.288   
## 2 2016 Aleutians West Census Area, Alaska ALASKA 0.236 0.288   
## 3 2050 Bethel Census Area, Alaska ALASKA 0.236 0.288   
## 4 15001 Hawaii County, Hawaii HAWAII 0.0638 0.103   
## 5 15003 Honolulu County, Hawaii HAWAII 0.0507 0.0848  
## 6 15005 Kalawao County, Hawaii HAWAII 0.061 0.0979  
## 7 30075 Powder River County, Montana MONTANA 0.264 0.313   
## 8 30083 Richland County, Montana MONTANA 0.264 0.313   
## 9 30093 Silver Bow County, Montana MONTANA 0.227 0.269   
## 10 30073 Pondera County, Montana MONTANA 0.267 0.323   
## # ... with 335 more rows, and 16 more variables:  
## # Estimated strongly hesitant <dbl>, Social Vulnerability Index (SVI) <dbl>,  
## # SVI Category <dbl>, CVAC level of concern for vaccination rollout <dbl>,  
## # CVAC Level Of Concern <dbl>,  
## # Percent adults fully vaccinated against COVID-19 (as of 6/10/21) <dbl>,  
## # Hispanic <dbl>, Percent non-Hispanic American Indian/Alaska Native <dbl>,  
## # Asian <dbl>, Black <dbl>, ...

##BAR Charts

B1 <- Northeast\_state\_data %>% ggplot(aes(x=`SVI Category`, fill=`SVI Category`)) + geom\_bar() +  
 labs(title= "Bar Chart of Frequency of Counties by SVI category (NOrtheast") + xlab("SVI CATEGORY") + ylab("Count") + theme\_classic()  
  
print(B1)

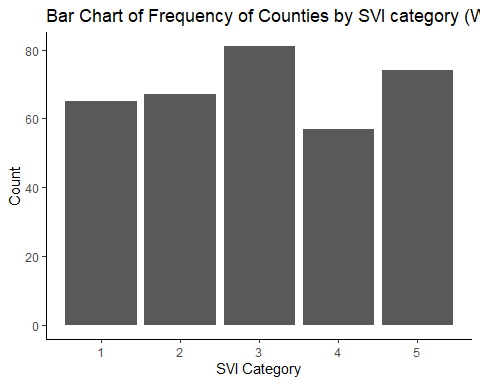


B2 <- Midwest\_state\_data %>% ggplot(aes(x=`SVI Category`, fill=`SVI Category`)) + geom\_bar() +  
 labs(title= "Bar Chart of Frequency of Counties by SVI category (MIDWEST)") + xlab("SVI Category") + ylab("Count") + theme\_classic()  
  
  
print(B2)

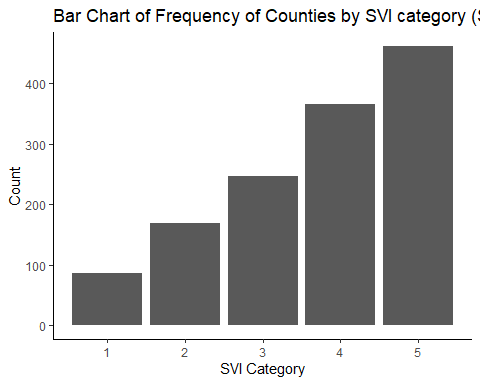


B3 <- West\_state\_data %>% ggplot(aes(x=`SVI Category`, fill=`SVI Category`)) + geom\_bar() +  
 labs(title= "Bar Chart of Frequency of Counties by SVI category (WEST)") + xlab("SVI Category") + ylab("Count") + theme\_classic()  
  
print(B3)

## Warning: Removed 1 rows containing non-finite values (stat\_count).



B4 <- South\_state\_data %>% ggplot(aes(x=`SVI Category`, fill=`SVI Category`)) + geom\_bar() +  
 labs(title= "Bar Chart of Frequency of Counties by SVI category (SOUTH)") + xlab("SVI Category") + ylab("Count") + theme\_classic()  
  
print(B4)



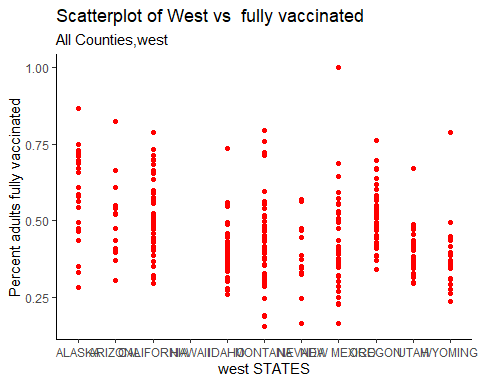
##SCatter plot between STATE and percent fully vaccinated

p1 <- West\_state\_data %>% ggplot(aes(x= State, y=`Percent adults fully vaccinated against COVID-19 (as of 6/10/21)`)) + geom\_point(color="red") + geom\_smooth(method='lm', color="black") + labs(title= "Scatterplot of West vs fully vaccinated", subtitle = ("All Counties,west")) + xlab("west STATES") + ylab("Percent adults fully vaccinated") + theme\_classic()  
  
print(p1)

## `geom\_smooth()` using formula 'y ~ x'

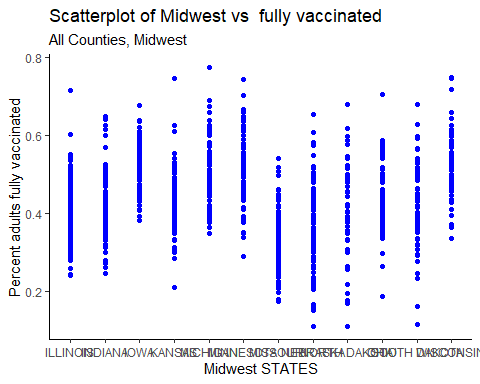
## Warning: Removed 15 rows containing non-finite values (stat\_smooth).

## Warning: Removed 15 rows containing missing values (geom\_point).



p2 <- Midwest\_state\_data %>% ggplot(aes(x= State, y=`Percent adults fully vaccinated against COVID-19 (as of 6/10/21)`)) + geom\_point(color="blue") + geom\_smooth(method='lm', color="black") + labs(title= "Scatterplot of Midwest vs fully vaccinated", subtitle = ("All Counties, Midwest")) + xlab("Midwest STATES") + ylab("Percent adults fully vaccinated") + theme\_classic()  
  
print(p2)

## `geom\_smooth()` using formula 'y ~ x'

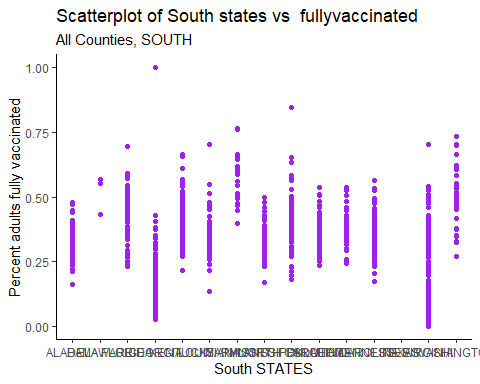


p3 <- South\_state\_data %>% ggplot(aes(x= State, y=`Percent adults fully vaccinated against COVID-19 (as of 6/10/21)`)) + geom\_point(color="purple") + geom\_smooth(method='lm', color="black") + labs(title= "Scatterplot of South states vs fullyvaccinated", subtitle = ("All Counties, SOUTH")) + xlab("South STATES") + ylab("Percent adults fully vaccinated") + theme\_classic()  
  
  
print(p3)

## `geom\_smooth()` using formula 'y ~ x'

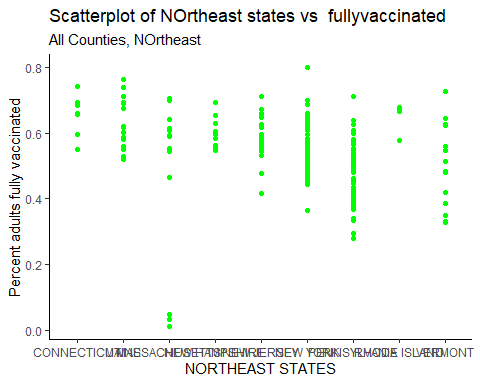
## Warning: Removed 263 rows containing non-finite values (stat\_smooth).

## Warning: Removed 263 rows containing missing values (geom\_point).



p4 <- Northeast\_state\_data %>% ggplot(aes(x= State, y=`Percent adults fully vaccinated against COVID-19 (as of 6/10/21)`)) + geom\_point(color="green") + geom\_smooth(method='lm', color="black") + labs(title= "Scatterplot of NOrtheast states vs fullyvaccinated", subtitle = ("All Counties, NOrtheast")) + xlab("NORTHEAST STATES") + ylab("Percent adults fully vaccinated") + theme\_classic()  
  
print(p4)

## `geom\_smooth()` using formula 'y ~ x'

 ## Full analysis

*Use one or several suitable statistical/machine learning methods to analyze your data and to produce meaningful figures, tables, etc. This might again be code that is best placed in one or several separate R scripts that need to be well documented. You want the code to produce figures and data ready for display as tables, and save those. Then you load them here.*

Example table 4.1 shows a table summarizing a linear model fit.

Table 4.1: Linear model fit table.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| term | estimate | std.error | statistic | p.value |
| (Intercept) | -43.7883068 | 61.1150617 | -0.7164896 | 0.4940713 |
| Height | 0.6996272 | 0.3675692 | 1.9033889 | 0.0934786 |

# 5 Discussion

## 5.1 Summary and Interpretation

*Summarize what you did, what you found and what it means.*

## 5.2 Strengths and Limitations

*Discuss what you perceive as strengths and limitations of your analysis.*

## 5.3 Conclusions

*What are the main take-home messages?*

*Include citations in your Rmd file using bibtex, the list of references will automatically be placed at the end*

This paper (Leek & Peng, 2015) discusses types of analyses.

Note that this cited reference will show up at the end of the document, the reference formatting is determined by the CSL file specified in the YAML header. Many more style files for almost any journal [are available](https://www.zotero.org/styles). You also specify the location of your bibtex reference file in the YAML. You can call your reference file anything you like, I just used the generic word references.bib but giving it a more descriptive name is probably better.

# References

Leek, J. T., & Peng, R. D. (2015). Statistics. What is the question? *Science (New York, N.Y.)*, *347*, 1314–1315. <https://doi.org/10.1126/science.aaa6146>