Gurpinder Singh

STAT 108

11/9/2022

The research questions is to infer if there is a connection between quality of school and home prices inside of California for the year 2021.

Load all the followin library

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.3.6      v purrr   0.3.4
## v tibble  3.1.8      v dplyr   1.0.10
## v tidyr   1.2.1      v stringr 1.4.1
## v readr   2.1.3      v forcats 0.5.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(stringr)
library(knitr)
library(skimr)
library(broom)
library(readr)
```

The following data tries to measure the quality of school. More specifically it takes into account the following variables: Absentness/Reason for absent, chronic Absentee, population, stability of student, suspension count, nation test results.

```
absentReason <- read.delim("data/schoolData/abreason2021.txt")
chronicAbsentee <- read.delim("data/schoolData/chrabs2021.txt")
cohort <- read.delim("data/schoolData/cohort2021.txt")
stabilityCount <- read.delim("data/schoolData/sr2021.txt")
suspeneded <- read.delim("data/schoolData/susp2021.txt")
test <- read.delim("data/schoolData/test/test.txt")
```

The response variable is housing price.

```
housing <- read_csv("data/housingData/housing.csv")
```

```
## Rows: 27424 Columns: 283
## -- Column specification --------------------------------------------------------
## Delimiter: ","
```

```
## chr   (7): RegionName, RegionType, StateName, State, City, Metro, CountyName
## dbl (276): RegionID, SizeRank, 2000-01-31, 2000-02-29, 2000-03-31, 2000-04-3...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

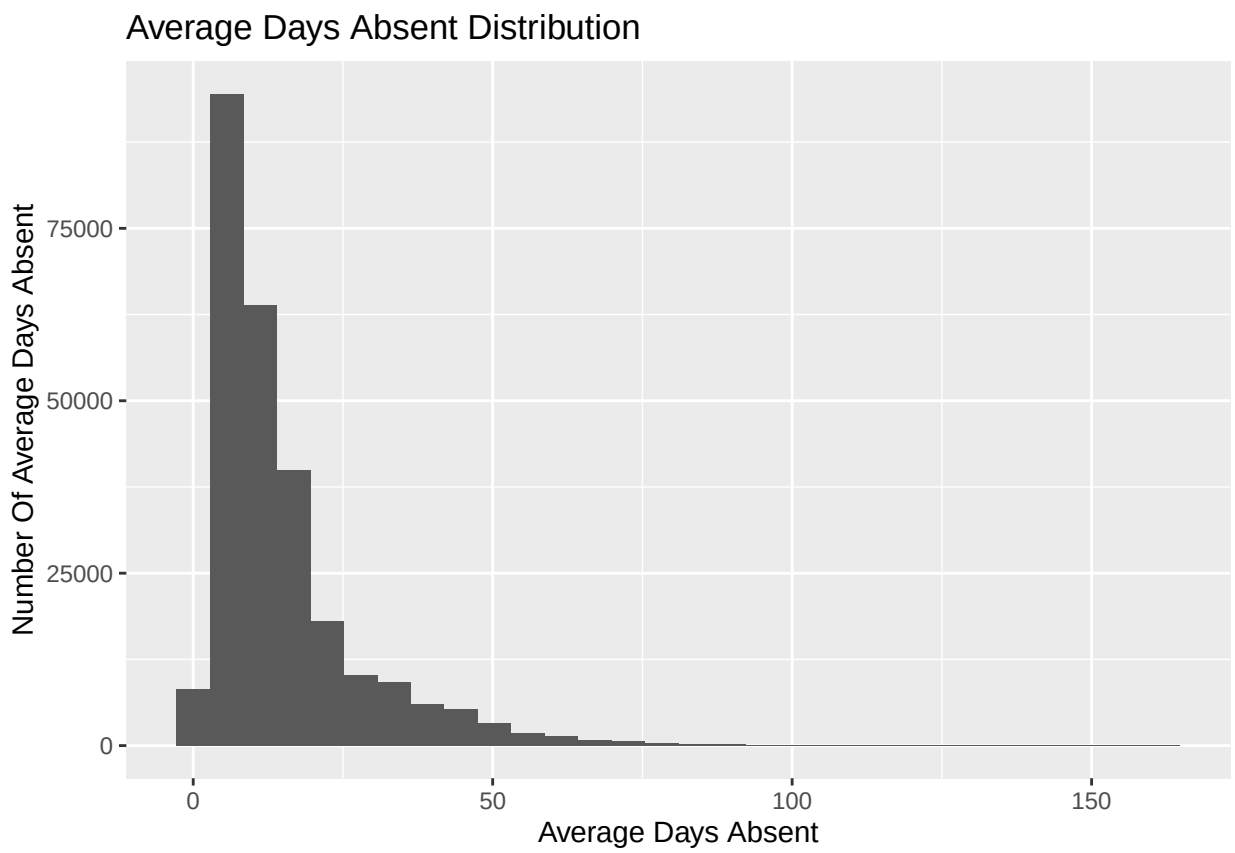
```
absentReason <- absentReason %>%
  mutate(Average.Days.Absent=round(as.numeric(absentReason$Average.Days.Absent), digits = 0))
```

```
## Warning in mask$eval_all_mutate(quo): NAs introduced by coercion
```

```
ggplot(data = absentReason, aes(x =Average.Days.Absent)) +
  geom_histogram() +
  labs(x ="Average Days Absent",
       y = "Number Of Average Days Absent",
       title = "Average Days Absent Distribution")
```

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

```
## Warning: Removed 107863 rows containing non-finite values (stat_bin).
```

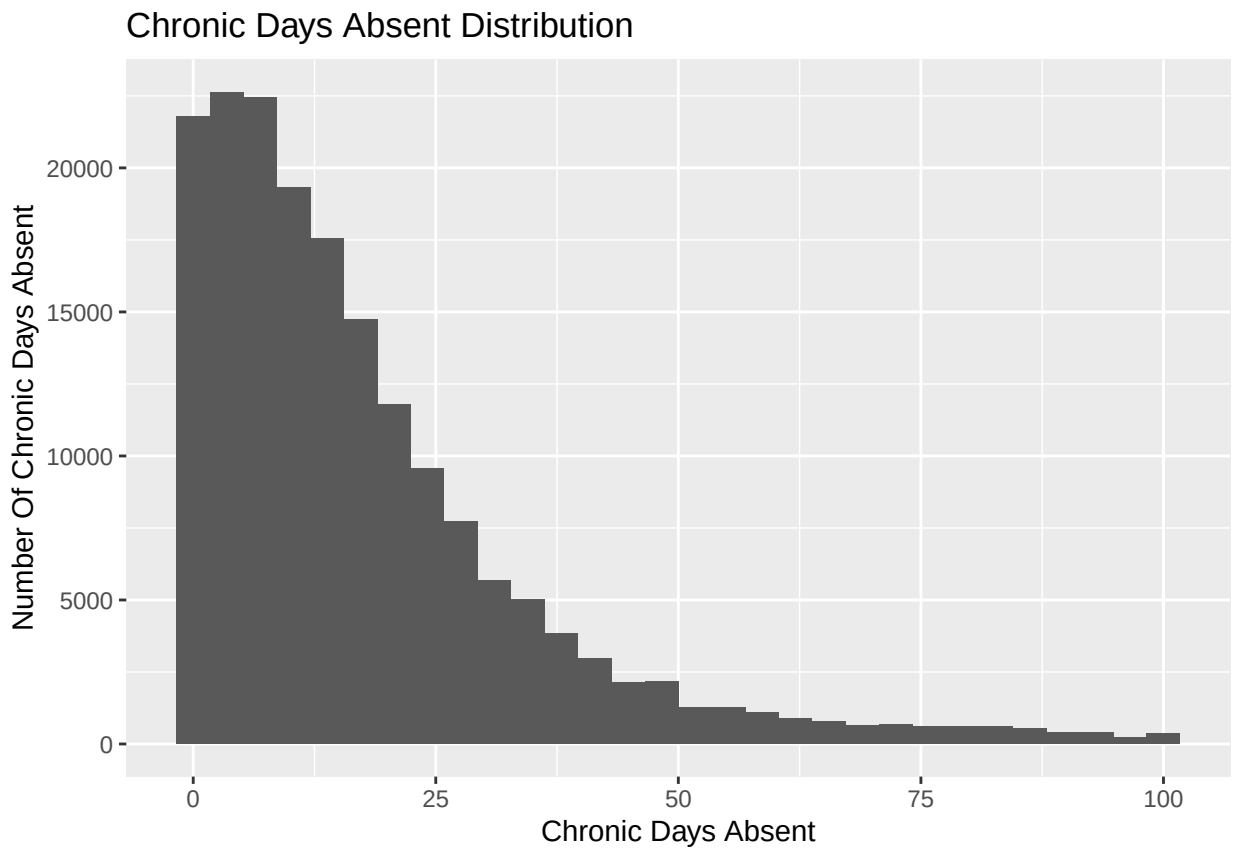



```
ggplot(data = chronicAbsentee, aes(x =ChronicAbsenteeismRate)) +
  geom_histogram() +
```

```
  labs(x ="Chronic Days Absent",
       y = "Number Of Chronic Days Absent",
       title = "Chronic Days Absent Distribution")
```

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

```
## Warning: Removed 83425 rows containing non-finite values (stat_bin).
```




```
glimpse(cohort)
```

```
## Rows: 254,938
## Columns: 34
## $ AcademicYear                  <chr> "2020-21", "2020-21", "2020-21~
## $ AggregateLevel                <chr> "C", "C", "C", "C", "C", "C", ~
## $ CountyCode                    <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ~
## $ DistrictCode                  <int> NA, NA, NA, NA, NA, NA, NA, NA~
## $ SchoolCode                    <int> NA, NA, NA, NA, NA, NA, NA, NA~
## $ CountyName                    <chr> "Alameda", "Alameda", "Alameda~
## $ DistrictName                  <chr> "", "", "", "", "", "", "", ""~
## $ SchoolName                    <chr> "", "", "", "", "", "", "", ""~
## $ CharterSchool                 <chr> "All", "All", "All", "All", "A~
## $ DASS                          <chr> "All", "All", "All", "All", "A~
## $ ReportingCategory             <chr> "GF", "GM", "GX", "RA", "RB", ~
## $ CohortStudents                <chr> "8683", "9158", "*", "4439", "~
```

```
## $ Regular.HS.Diploma.Graduates..Count.    <chr> "7821", "7678", "*", "4237", "~
## $ Regular.HS.Diploma.Graduates..Rate.     <chr> "90.1", "83.8", "*", "95.4", "~
## $ Met.UC.CSU.Grad.Req.s..Count.           <chr> "5115", "4416", "*", "3316", "~
## $ Met.UC.CSU.Grad.Req.s..Rate.            <chr> "65.4", "57.5", "*", "78.3", "~
## $ Seal.of.Biliteracy..Count.              <chr> "1494", "859", "*", "1074", "2~
## $ Seal.of.Biliteracy..Rate.               <chr> "19.1", "11.2", "*", "25.3", "~
## $ Golden.State.Seal.Merit.Diploma..Count. <chr> "2571", "1970", "*", "2228", "~
## $ Golden.State.Seal.Merit.Diploma..Rate   <chr> "32.9", "25.7", "*", "52.6", "~
## $ CHSPE.Completer..Count.                 <chr> "36", "25", "*", "22", "3", "0~
## $ CHSPE.Completer..Rate.                  <chr> "0.4", "0.3", "*", "0.5", "0.2~
## $ Adult.Ed..HS.Diploma..Count.            <chr> "0", "3", "*", "0", "0", "0", ~
## $ Adult.Ed..HS.Diploma..Rate.             <chr> "0.0", "0.0", "*", "0.0", "0.0~
## $ SPED.Certificate..Count.                <chr> "57", "97", "*", "18", "27", "~
## $ SPED.Certificate..Rate.                 <chr> "0.7", "1.1", "*", "0.4", "1.6~
## $ GED.Completer..Count.                   <chr> "0", "5", "*", "1", "0", "0", ~
## $ GED.Completer..Rate.                    <chr> "0.0", "0.1", "*", "0.0", "0.0~
## $ Other.Transfer..Count.                  <chr> "15", "29", "*", "2", "3", "1"~
## $ Other.Transfer..Rate.                   <chr> "0.2", "0.3", "*", "0.0", "0.2~
## $ Dropout..Count.                         <chr> "449", "849", "*", "112", "172~
## $ Dropout..Rate.                          <chr> "5.2", "9.3", "*", "2.5", "10.~
## $ Still.Enrolled..Count.                  <chr> "305", "472", "*", "47", "139"~
## $ Still.Enrolled..Rate.                   <chr> "3.5", "5.2", "*", "1.1", "8.1~
```