# Untitled

Shan

11/17/2020

## Contents



```
knitr::opts_chunk$set(echo = FALSE, fig.width = 7, fig.align = 'center', fig.asp = 0.618, out.width = "
library(data.table)
library(tidyverse)
```

```
## -- Attaching packages ---------------------------------------------------------------------------
```

```
## v ggplot2 3.3.2     v purrr   0.3.4
## v tibble  3.0.3     v dplyr   1.0.2
## v tidyr   1.1.1     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.5.0
```

```
## -- Conflicts ------------------------------------------------------------------------------------
## x dplyr::between()   masks data.table::between()
## x dplyr::filter()    masks stats::filter()
## x dplyr::first()     masks data.table::first()
## x dplyr::lag()       masks stats::lag()
## x dplyr::last()      masks data.table::last()
## x purrr::transpose() masks data.table::transpose()
```

```
library(psych)
```

```
##
## Attaching package: 'psych'
```

```
## The following objects are masked from 'package:ggplot2':
##
##     %+%, alpha
```

```
library(leaflet)
library(ggplot2)
library(dplyr)
```

Data Souces: 1.2019 races number data:https://www.kff.org/other/state-indicator/distribution-by-raceethnicity/?dataView=1&currentTimeframe=0&selectedRows=%7B%22wrapups%22:%7B%22united-states%22:%7B%7D%7D%7D&sortModel=%7B%22colId%22:%22Location%22,%22sort%22:%22asc%22%7D 2.Covid data:https://covidtracking.com/race/about#download-the-data

-Introduction: Key Question: Among population with known race category across the U.S, which race was being affected most?

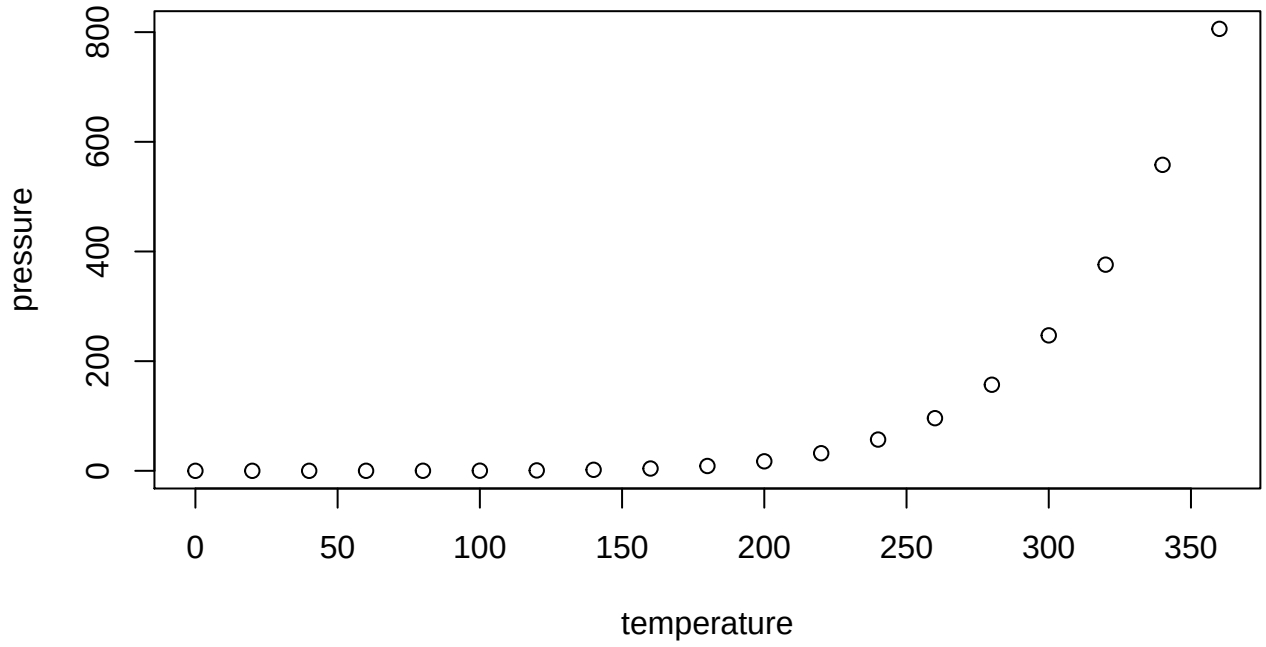
data background: The primary data source is the COVID Racial Data Tracker.They collect, cross-check, and publish COVID-19 data from 56 US states and territories in three main areas: testing, patient outcomes, and, via The COVID Racial Data Tracker, racial and ethnic demographic information. Data is being updated twice per week.I also combined the KFF data:Population Distribution by Race/Ethnicity(2019) for the first and second questions.The KFF.org data is a non-profit foundation that does research, journalism, and communications programs.

-Methods: Given the data above, we consider removing the NA's. We also excluded the "other" category in COVID data since there is no correspond variable in the population data. We sum up the case number of each race by the latest updated numbers, then look at the actual number of cases and deaths, prevalence, death rate and case fatality rate. The software using R. Prevalence: the proportion of a population who have a specific characteristic in a given time period. Therefore, we use case number divided by the total number of population to get the prevalence for each known races.The time period would be from 04/12/2020 to 08/11/2020 for all the calculated rate. Death rate: a measure of the number of deaths (in general, or due to a specific cause) in a particular population, scaled to the size of that population, per unit of time. We use death number devided by the total number of population to get the death rate for each know race. Case fatality rate (CFR) : the proportion of deaths from a certain disease compared to the total number of people diagnosed with the disease for a particular period.So we use de

```
##      speed           dist       
##  Min.   : 4.0   Min.   :  2.00  
##  1st Qu.:12.0   1st Qu.: 26.00  
##  Median :15.0   Median : 36.00  
##  Mean   :15.4   Mean   : 42.98  
##  3rd Qu.:19.0   3rd Qu.: 56.00  
##  Max.   :25.0   Max.   :120.00
```

## Including Plots

You can also embed plots, for example:




Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.