Due: 22nd January 2024, 6pm

1. Using the R-help explain what each of the commands mean:
   (a) `read.csv`
   (b) `setwd`
   (c) `getwd`
2. Download data set dice.csv from moodle. `Create a folder/directory named Dice` and place the data set there. Using `setwd`, or `Session` command to set your working directory to Dice.
   (a) Run the following `R-code`:

```
DiceR = read.csv("Dice.csv", header=T)
names(DiceR)= "Sum"
DiceR$Sum = as.numeric(DiceR$Sum)
Dice = na.omit(DiceR$Sum)
hist(Dice)
summary(Dice)
```

   Explain each line in the code above and attach the output(s) to the answer.
   (b) Run the following `R-code`:

```
 cs = (Dice-mean(Dice))/(sd(Dice))
mean(cs)
sd(cs)
onesdcs = cs[cs > -1& cs <1]
twosdcs = cs[cs > -2& cs <2]
threesdcs = cs[cs > -3& cs <3]
```

   Explain each line in the code above. Can you decide if the sum of 5 rolls are approximated well by a Normal distribution ? Explain and justify your answer.
3. Download data set from: https://data.incovid19.org/csv/latest/states.csv.
   (a) Go to https://data.incovid19.org and write out a one paragraph description of what the data set contains. Load `state.csv` in R into a dataframe called `statedf`.
   (b) Pick a state of India which has the same starting letter as the second letter in your last name. For E.g.: Siva Athreya could pick Tamil Nadu[1].
   (c) Subset the dataframe `statedf` to have only data from the state that you picked in the previous step and call the resulting dataframe as `mystatedf`

[1] If there is no such state then use the second letter and proceed till you are able to find a state

(d) Using `mystatedf` compute the daily active cases for the state. Then plot a line chart using `geom_line` for the period July 1, 2021 to December 31, 2021,`viridis` colored by date. Provide: Title as "Active cases for State– NAME"; Xlabel - dates, Ylabel - Cases, X-ticks to be dates.

(e) Using `mystatedf` compute the total `Deceased` figures for each month since March 2020 till December 31, 2021. Then plot a bar chart using `geom_bar` of the monthly Deceased totals, `viridis` colored by month. Provide: Title as "Monthly Deceased Totals for State– NAME"; Xlabel - Months, Ylabel - Deceased Total, X-ticks to be names of months.

(f) Using `mystatedf` compute the total `Confirmed` cases for each month from March 2020 till December 31, 2021. Then provide a box plot of the same.

(g) Using `statedf` compute the total `Confirmed` cases and total `Deceased` for each state since March 2020 till December 31, 2021. Then plot a scatter using `geom_point` of the total confirmed cases of the states versus total deceased figures; `viridis` colored by state. Provide: Title as "Scatter Plot of Confirmed Versus Deceased"; Xlabel - Deceased Figures, Ylabel - Confirmed Cases. *Can you label the dots with the State names ?*

Upload the worksheet as one pdf file on moodle that provides all answers, graphs as images along with `R-code` as applicable.