Homework 2

MTH 3270

Tobias Boggess

2/7/2022

Appendix B Problems

Question 1: Do the following

a) Write R commands that create three vectors:

- TrueAndMissing containing values TRUE and NA (at least one of each in any order).
- FalseAndMissing containing values FALSE and NA.
- Mixed containing values TRUE, FALSE and NA.

Report your R commands.

```
TrueAndMissing <- c(T, T, NA, T, T, NA)
TrueAndMissing
```

```
## [1] TRUE TRUE   NA TRUE TRUE   NA
```

```
FalseAndMissing <- c(F, NA, F, F, NA, F)
FalseAndMissing
```

```
## [1] FALSE    NA FALSE FALSE    NA FALSE
```

```
Mixed <- c(T, F, NA, T, F, T, T, F, NA)
Mixed
```

```
## [1]  TRUE FALSE    NA  TRUE FALSE  TRUE  TRUE FALSE    NA
```

b) Apply the function any() and all() to each of the vectors of part *a* and report the results.
Results:

```
# using the any() function for the vectors in part a
any(TrueAndMissing)
```

```
## [1] TRUE
```

```
any(FalseAndMissing)
```

```
## [1] NA
```

```
any(Mixed)
```

```
## [1] TRUE
```

```
# using the all() function for the vectors in part a
all(TrueAndMissing)
```

```
## [1] NA
```

```
all(FalseAndMissing)
```

```
## [1] FALSE
```

```
all(Mixed)
```

```
## [1] FALSE
```

Question 2: Do the following with the following datasets.

Datasets:

```
# Illiteracy rates per state
illit <- c(2.1, 1.5, 1.8, 1.9, 1.1, 0.7, 1.1, 0.9, 1.3, 2.0)

# Murder rates per state
murder <- c(15.1, 11.3, 7.8, 10.1, 10.3, 6.8, 3.1, 6.2, 10.7, 13.9)

# Population per state
pop <- c(3615, 365, 2212, 2110, 21198, 2541, 3100, 579, 8277, 4931)

# States used for data
state <- c("Alabama", "Alaska", "Arizona", "Arkansas", "California",
"Colorado", "Connecticut", "Delaware", "Florida", "Georgia")
```

a) Return the names of the states whose populations are less than 2,500 (thousand)
Results:

```
# Returns the states with a population under 2500
state[pop < 2500]
```

```
## [1] "Alaska"   "Arizona"  "Arkansas" "Delaware"
```

b) Return illiteracy rates that are greater than the median illiteracy rate
Results:

```
# Returns illiteracy rates greater than the median
illit[illit > median(illit)]
```

```
## [1] 2.1 1.5 1.8 1.9 2.0
```

c) Return the murder rates for states whose illiteracy rate is greater than the median illiteracy rate
Results:

```
# Returns the murder rates for states with greater illiteracy rates than the median
murder[illit > median(illit)]
```

```
## [1] 15.1 11.3  7.8 10.1 13.9
```

Question 3: Do the following with the data set from question 2

a) Write a command involving which() that determines the indices of the murder rates that are greater than 12.
Results:

```
# Returns indices of murder rates greater than 12
which(murder > 12)
```

```
## [1]  1 10
```

b) Why does the expression "(1:10)[murder > 12]" contain parenthesis
Results:

```
which(murder > 12)
```

```
## [1]  1 10
```

```
(1:10)[murder > 12]
```

```
## [1]  1 10
```

The parenthesis tell R how many elements to consider in the murder vector and then it will print all the elements in the vector that meet the requirements.If the term in the parenthesis is 1:5 then it will only consider the first 5 elements and if the parameter is 1:12 then it will consider 11 elements even though there are only ten elements in the murder vector.

Question 4: Guess whether the following will evaluate to TRUE or FALSE

a) $(4 > 5 \ \& \ 8 < 9) \mid 2 > 3$
My Guess: FALSE
Results:

```
 (4 > 5 & 8 < 9) | 2 > 3
```

```
## [1] FALSE
```

b) $4 > 5$ & $(8 < 9 \mid 2 > 3)$
My Guess: FALSE
Results:

```
4 > 5 & (8 < 9 | 2 > 3)
```

```
## [1] FALSE
```

c) $(4 > 5$ & $2 > 3) \mid 8 < 9$
My Guess: TRUE
Results:

```
(4 > 5 & 2 > 3) | 8 < 9
```

```
## [1] TRUE
```

d) $(4 > 5$ & $(2 > 3 \mid 8 < 9)) \mid 7 < 6$
My Guess: FALSE
Results:

```
(4 > 5 & (2 > 3 | 8 < 9)) | 7 < 6
```

```
## [1] FALSE
```

Question 5: Do the following. Explain the results.

a) Values for below.
Results:

```
NA | TRUE
```

```
## [1] TRUE
```

```
FALSE & NA
```

```
## [1] FALSE
```

NA | TRUE will always be true because the or operator and the logical value for TRUE will always equate to TRUE. Anything with FALSE and a logical and will equate the expression as FALSE.

b) Values for below.
Results:

```
NA | FALSE
```

```
## [1] NA
```

```
TRUE & NA
```

```
## [1] NA
```

Both of the expressions will report as NA because NA | FALSE is unknown for whether the expression will equate TRUE or FALSE. The TRUE & NA expression reports as NA because the expression doesn't know the value for NA so the expression cannot equate to either TRUE or FALSE.

Question 6: Write commands to return elements that don't include NA in the given vector.

Vector:

```
x <- c(12, 4, 8, NA, 9, NA, 7, 12, 13, NA, 10)

# Command that displays any values that don't include NA's
x[!is.na(x)]
```

```
## [1] 12  4  8  9  7 12 13 10
```

Question 7: Write a function for the geometric mean.

R Code with test cases:

```
gm <- function(x) {
  if(!is.numeric(x) | any(x <= 0) | any(is.na(x))) {
    stop("Negative or non-numeric values found")
  }

  geoMean <- prod(x)^(1/length(x))
  geoMean
}

# Test with all positive values
y <- c(12, 4, 18, 202, 4, 63, 456)
try(gm(y))
```

```
## [1] 29.63106
```

```r
# Test with positive and negative values
z <- c(45, 57, 67, 1, 20, 45, -56, -48, 12, -75)
try(gm(z))
```

```
## Error in gm(z) : Negative or non-numeric values found
```

```r
# Test with NA values
f <- c(45, 75, 81, NA, 64, 75, 81, 91, 2, NA, 45, NA)
try(gm(f))
```

```
## Error in gm(f) : Negative or non-numeric values found
```

```r
# Test with a character vector
g <- c("Bruce", "Amy", "John", "TRUE", "FALSE")
try(gm(g))
```

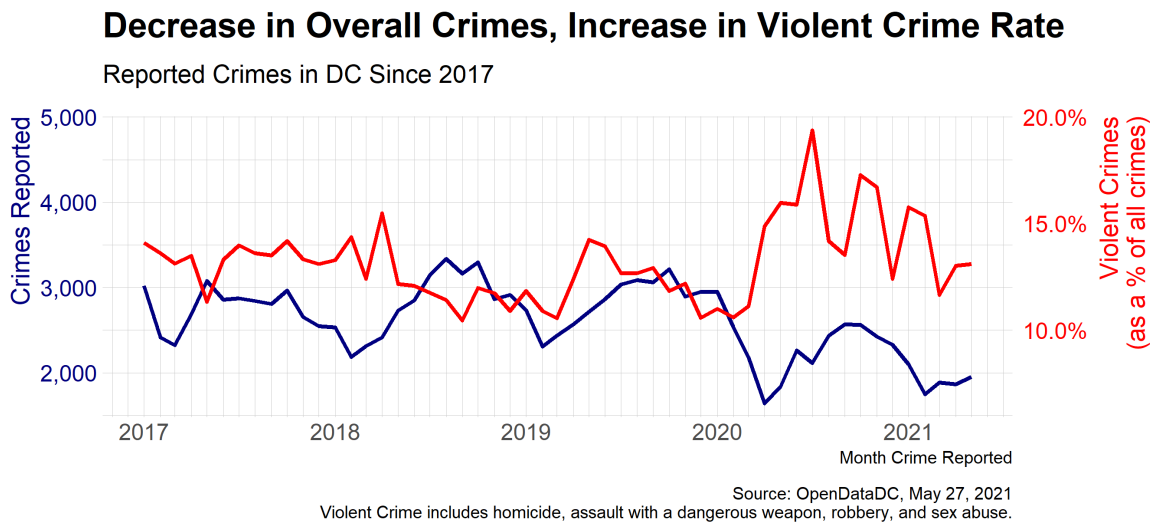
```
## Error in gm(g) : Negative or non-numeric values found
```

Chapter 2 Problems

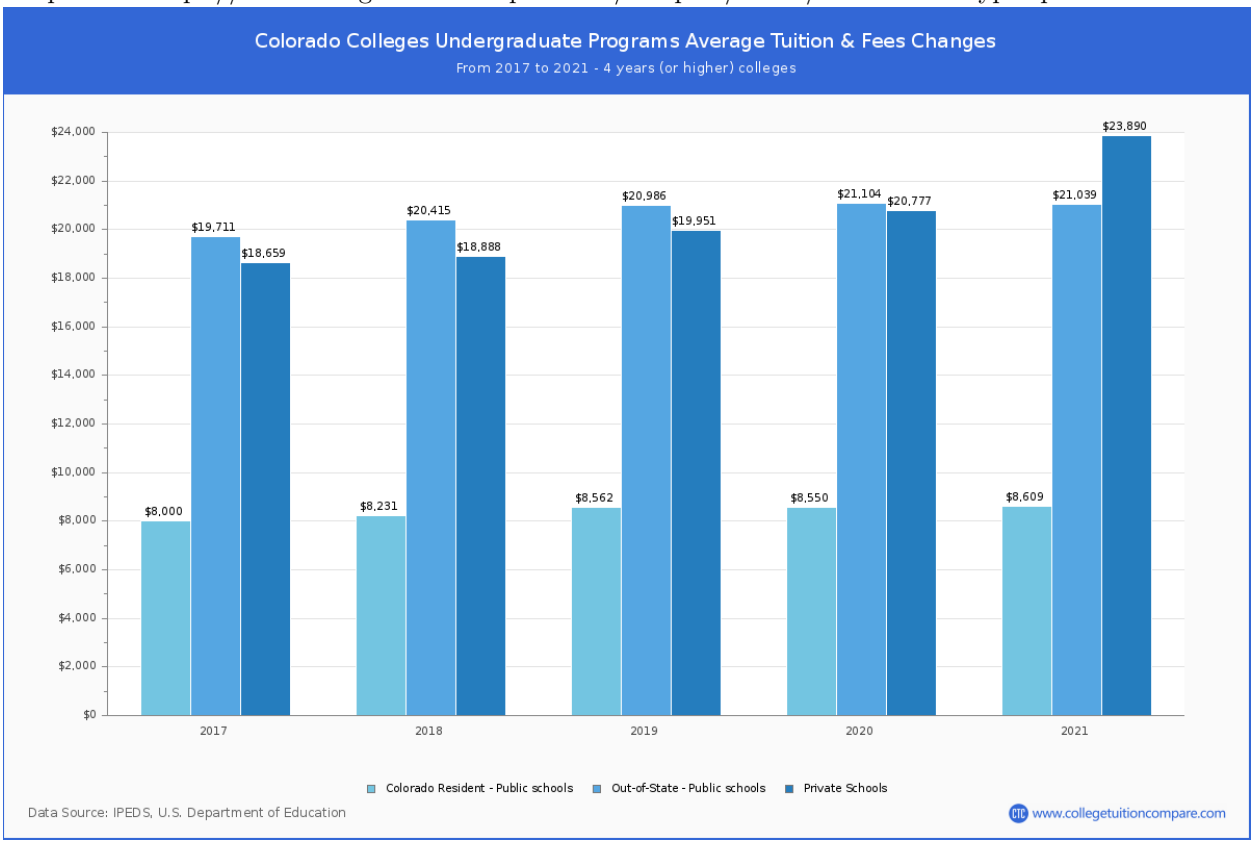
Question 3: Answer the following about two graphs found on the internet.

Graphs:

Graph from https://hershgupta.com/2021/05/27/dc-crime-rates/



Graph from https://www.collegetuitioncompare.com/compare/tables/?state=CO&type=public



- a) What aspects of the display work well, and how do these relate to the principles established in this chapter?

  The crime rate for the DC area graph does a good job with showing the direction of the crimes reported and the violent crimes percentage. Another well done thing is the different color patterns used. Furthermore, the graph shows the source and what crimes are included in the data. The scale of the graph is easy to read since the graph has numerical versus linear time scales. Another well done aspect of the graph tells the reader exactly what the data represents from the title, the x-axis, and the y-axis.

  The graph displaying the data, "average tuition and fees change", does a good job displaying the area of the bars from different years to clearly demonstrate the differences. Another quality feature is that the actual numbers are included right above each bar. Another great feature of the bar graph is the numerical versus time scales.

- b) What aspects of the display don't work well? Are there ways that the display might be improved?

  The first graph (overall crime graph) would be hard to read if someone was colorblind because of the red and the dark blue. Another possible fix could be to separate the graphs since there's a lot of overlap between the violent crime percentages and the crimes reported. Personally, I would prefer if the violent crimes were the number of violent crimes reported instead of percentages since that would be consistent with the crimes reported.

  The Colorado college tuition and fee changes graph could change a few features. One of which is the shades of blues of each bar in the graph because they are too similar and the graph would be easier to read if they were different colors other than different shades. Another thing to change would be the legend because it's a little small. It would be nice if it was larger so people who have trouble seeing small text could easily read the information.