Homework 1 - Zak Gilliam

Part 1

Zak
Gilliam
Written

Bios
4120

HW-1

1. Ordinal data has a specific ordering. Nominal data has no numerical relationship between values
2. Relative frequency would be useful when comparing two datasets of different sizes
3. a. discrete
   b. ordinal
   c. binary
   d. continuous
   e. binary
   f. ordinal
   g. discrete
   h. nominal
4. a. $\frac{1}{n}\sum_{i=1}^{13} x_i \approx \frac{1}{13}\sum_{i=1}^{13} x_i \approx \boxed{25.6}$ months
   b. $\boxed{120}$ months
   c. $[0.1, 96]$
   d. $IQR = Q_3 - Q_1$
      $= 38 - 41$
      $= \boxed{34}$ months
5. a. mean $\frac{1}{5}\sum_{n=1}^{5}(x_n) = \boxed{12.94}$ mm Range = $[9.2, 15.1]$
   median : $\boxed{11.4}$ mm
   std = $\sqrt{\frac{1}{5-1}\sum_{n=1}^{5}(x_n - 12.94)^2} = 3.6$

   b $\sum_{n=1}^{5}(x_n - 12.94) = (9.2-\bar{x}) + (10.9-\bar{x}) + (11.4-\bar{x}) + (15.1-\bar{x}) + (15.1-\bar{x})$
   $= -3.74 + -2.04 + -1.54 + 2.16 + 5.16$
   $= 0$

   c. mean $= \frac{1}{6}\sum_{i=1}^{6} x_i = 21.95$ // median $= 22.05$
   stdev $= \sqrt{\frac{1}{6-1}\sum_{n=1}^{6}(x_n - 21.95)^2} = 7.14$ // range $= [16.7, 32.5]$

Zak
Gilliam

Bios
4120

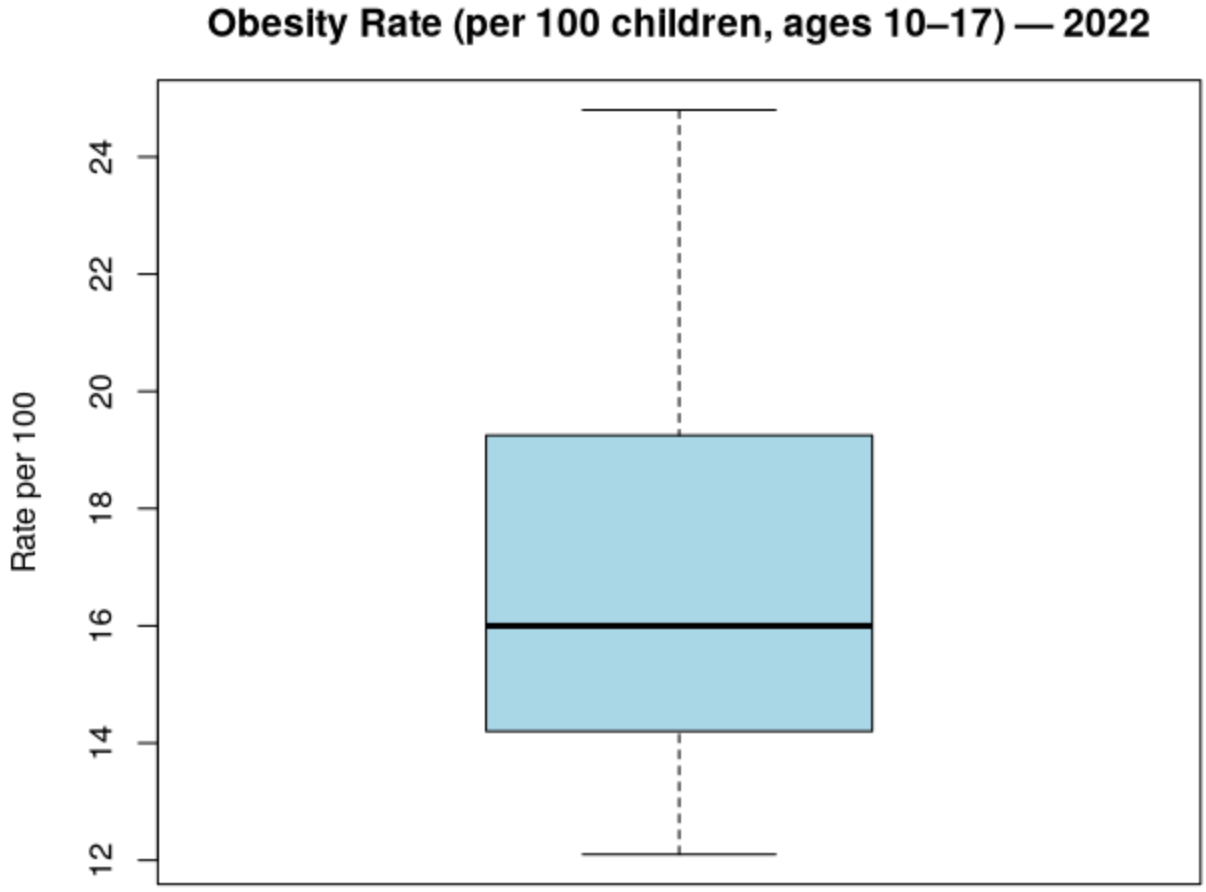
HW-1 Cont

6. Based on the data presented, no. While there is a positive slope in deaths overtime, there is no data presented that represents Public health. Additional variables, like % of population with chronic disease, access to sanitation, etc. would be necessary to prove this relationship.

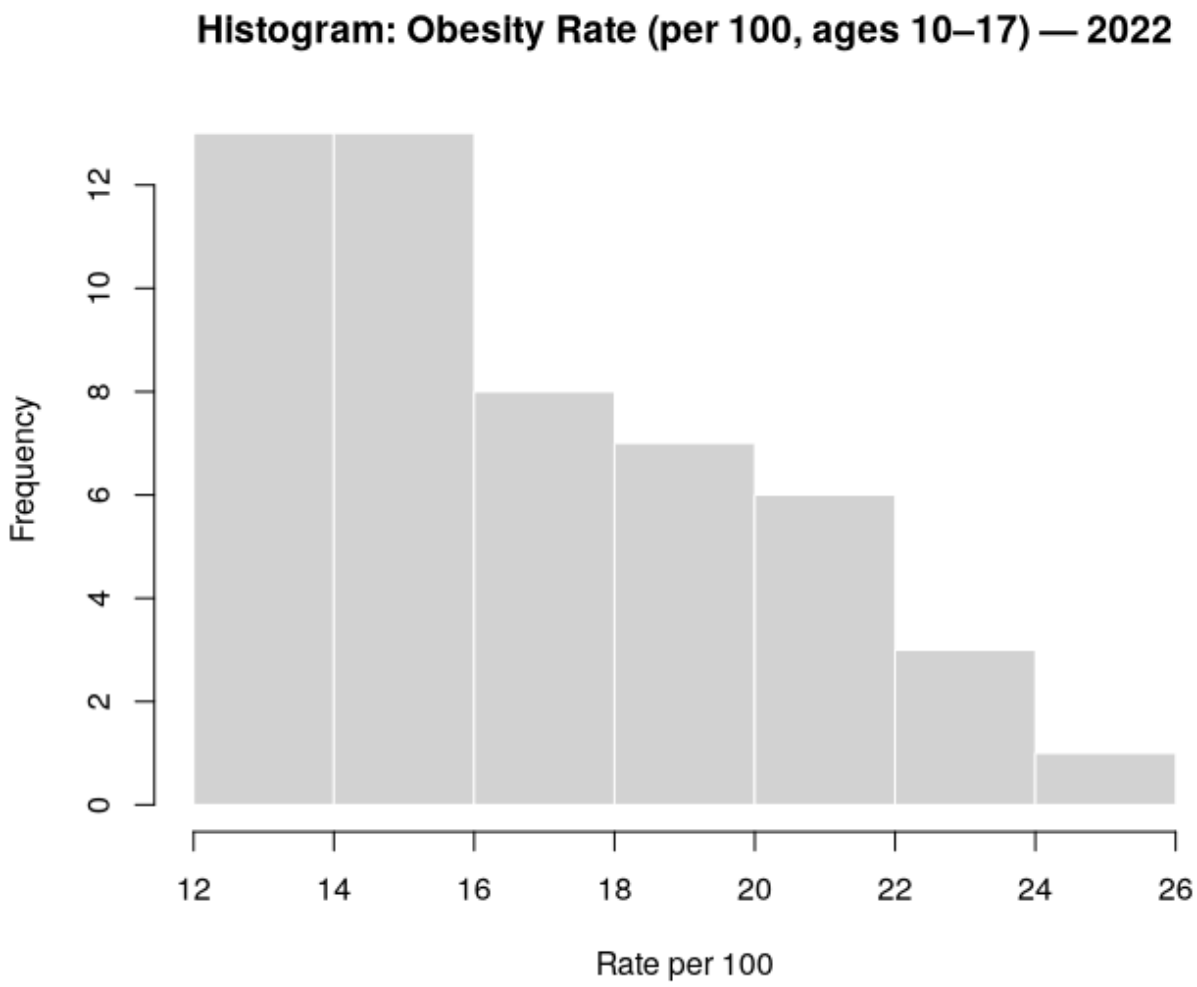
Part 2

Question 7

- Lowest rate: 12.1 in Minnesota, Highest rate: 24.8 in West Virginia, socioeconomic differences, school/health policies, healthcare access, cultural/dietary patterns, and demographics.



-
- The data is right skewed, no states are outliers via the IQR rule



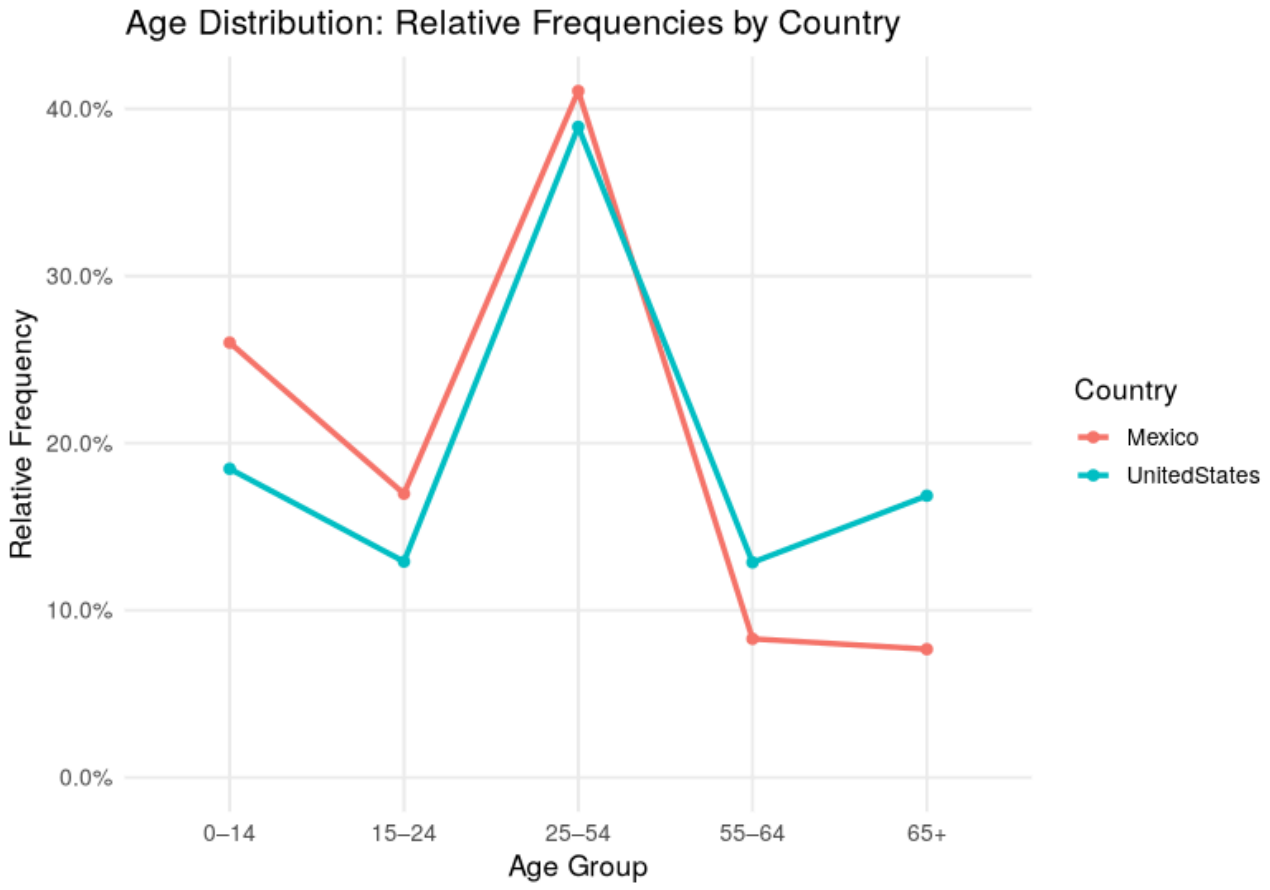
- I find this more informative. I believe histograms convey information regarding the shape of a distribution more effectively than box plots.
- As a summary of the 'average state', the simple mean is ok. - For a national prevalence, it is NOT ideal because states have different population sizes. A population-weighted mean (weighted by each state's 10–17 population) is more appropriate.

Question 8

- No, it is not reasonable. There is a considerable difference it total population between the countries for each age group.
-

	AgeGroup	Mexico	UnitedStates	MexicoRel	USRel
1	0–14	33461.0	61408.9	0.2601	0.1846
2	15–24	21832.0	42937.8	0.1697	0.1291
3	25–54	52827.9	129458.2	0.4106	0.3892
4	55–64	10663.2	42782.5	0.0829	0.1286
5	65+	9865.4	56051.6	0.0767	0.1685
6	Total	128649.5	332639.0	1.0000	1.0000

`MexicoRel` and `USRel` are the relative frequencies for Mexico and the US respectively

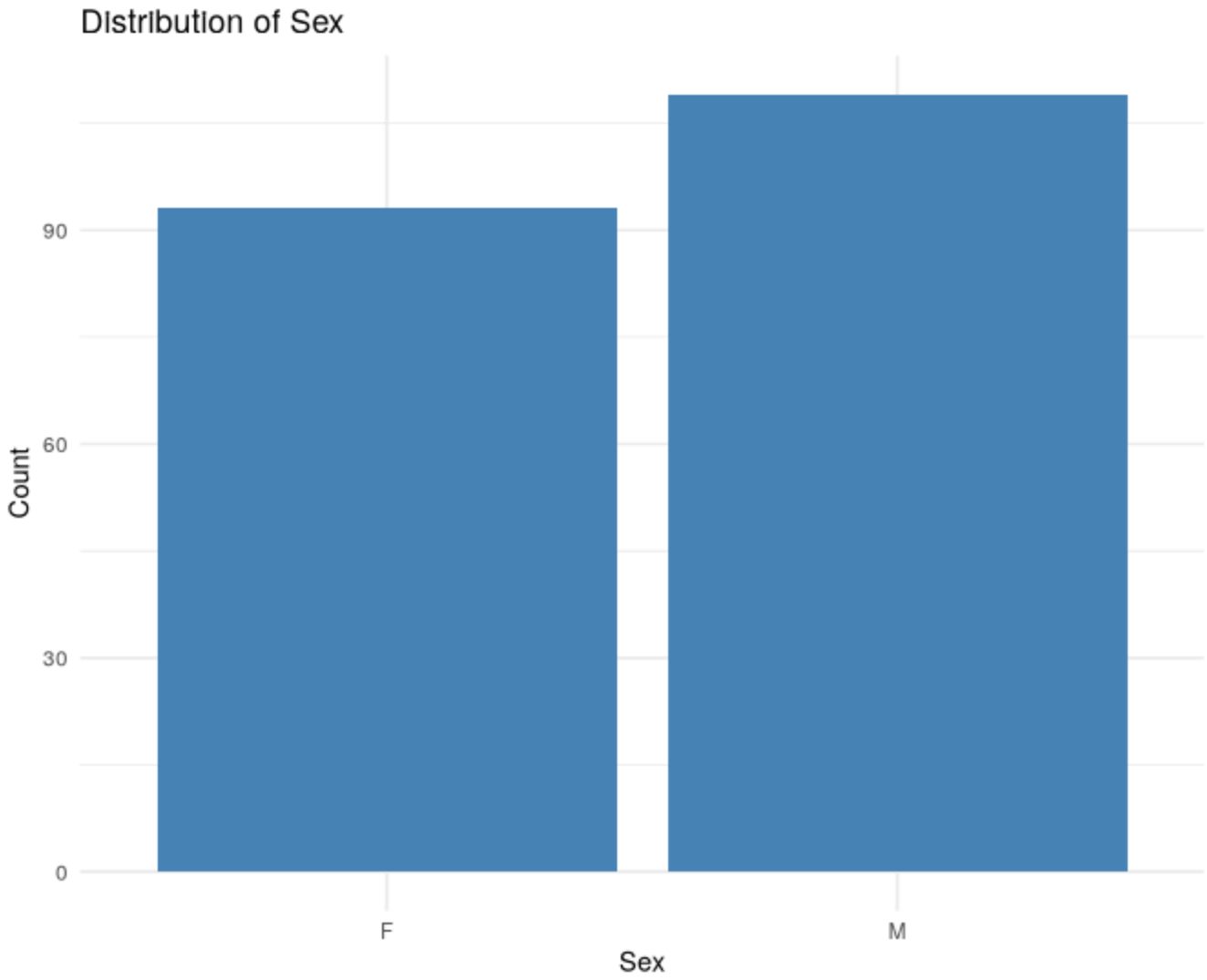


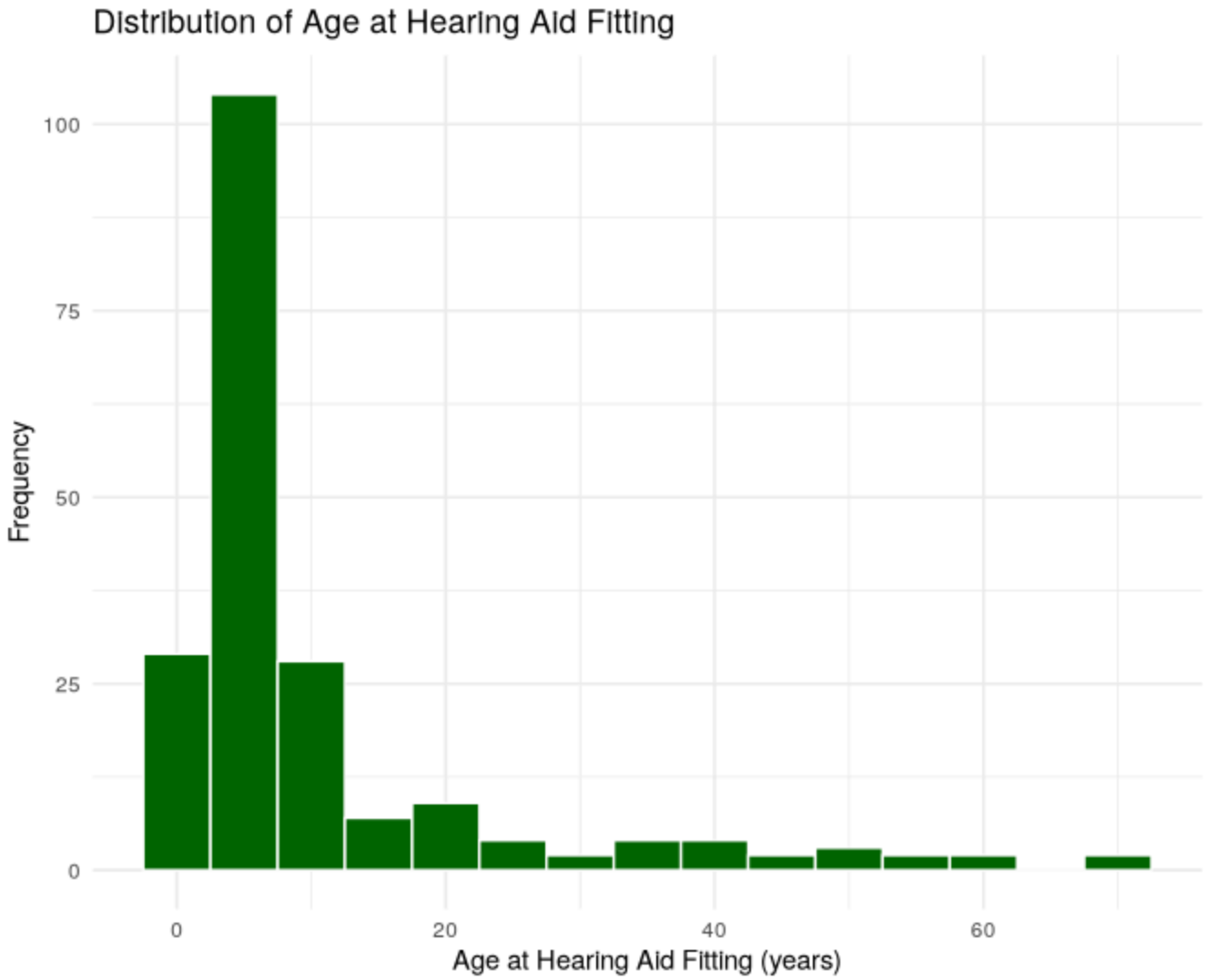
- Mexico shows a younger population structure with higher shares under 25, while the U.S. has relatively more people in older age groups (55+).

Question 9

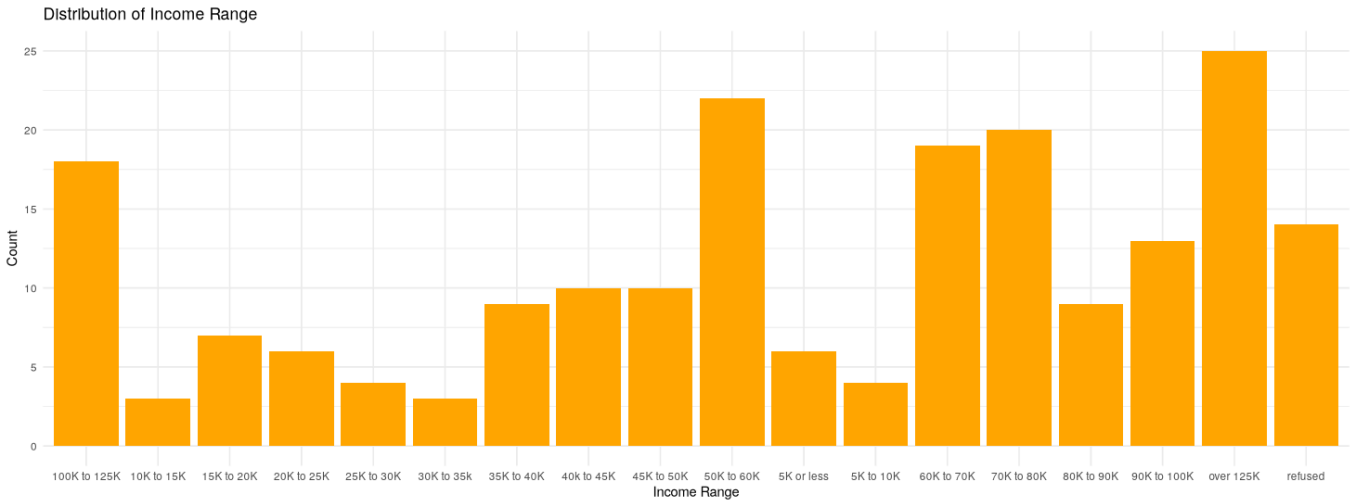
A: Categorical - Nominal
B: Sex - Binary (categorical/nominal)
C: Age aid fitted - continuous, 21, min = 2, max = 72
D: Income Range - categorical, ordinal
E: better ear pta - continuous, dB HL (decibels hearing level), ~15.0 to 82.5 (represents degree of hearing loss, lower = better hearing, higher = worse)

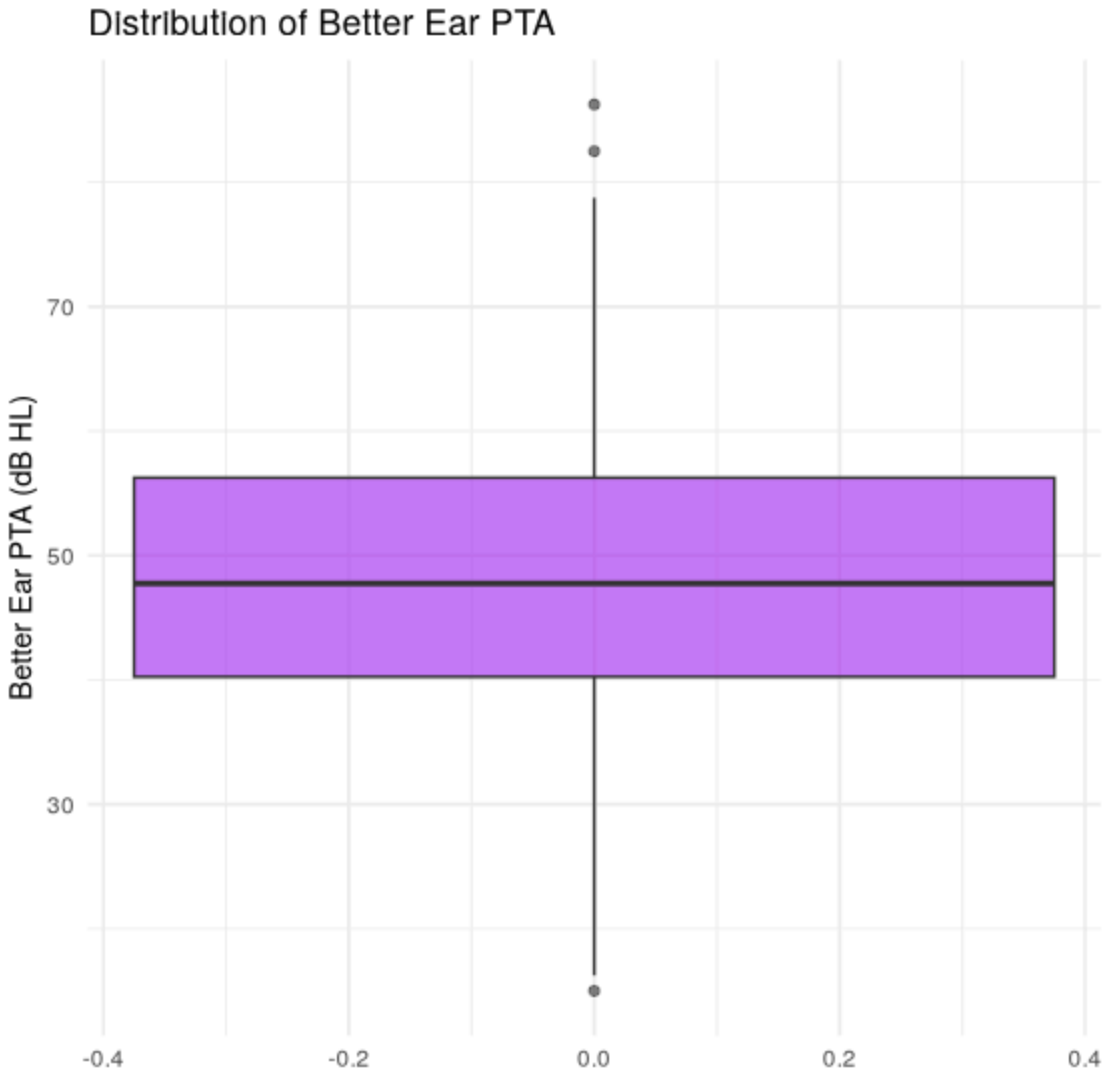
Question 10


Distribution of Sex
Count
90
60
30
0
F
M
Sex



Data is right skewed with upper outliers


Distribution of Income Range
Count
25
20
15
10
5
0
100K to 125K
10K to 15K
15K to 20K
20K to 25K
25K to 30K
30K to 35k
35K to 40K
40k to 45K
45K to 50K
50K to 60K
5K or less
5K to 10K
60K to 70K
70K to 80K
80K to 90K
90K to 100K
over 125K
refused
Income Range


Distribution of Better Ear PTA
Better Ear PTA (dB HL)
70
50
30
-0.4
-0.2
0.0
0.2
0.4

Mostly normal distribution with two upper and one lower outliers