Lab 1, Part 1

Section 14, Group 2: Ethan Duncan, Jeremy Lan, Nicolas Loffreda

1.1 Professional magic

1.1.1 What is the type I error of the test?

A type I error occurs when the null hypothesis is true but we reject it. Its associated probability is $\alpha = P(\text{reject } H_0 | H_0 \text{ is true})$.

In this case, we assume that $H_0 : p = \frac{1}{2}$ and we set the rejection region when the test statistic $S = X_1 + Y_1 + X_2 + Y_2 + X_3 + Y_3 \in \{0, 6\}$. So, the type I error of the test is the probability that S takes the values 0 or 6, assuming that $p = 1/2$, or $P(S = 0 \cup S = 6 | p = \frac{1}{2})$.

We first need to notice that $S = 0$ or $S = 6$ will happen only when all X_i and Y_i are the same. All variables need to take the value 0 or all the variables need to be 1. From the joint distribution, we can see that $P(X_i = Y_i = 0) = P(X_i = Y_i = 1) = p/2$. The probability of either of these things for happening is just the sum of both, so $P(X_i = Y_i) = p$.

So, $P(S = 0 \cup S = 6)$ can be written as:

$$P(X_1 = Y_1 \cap X_2 = Y_2 \cap X_3 = Y_3)$$

Because we know that each pair is independent of each other:

$$P(X_1 = Y_1) \cdot P(X_2 = Y_2) \cdot P(X_3 = Y_3) = p \cdot p \cdot p = p^3$$

Finally, to calculate α we need to assume that H_0 is true:

$$\boxed{\alpha = P(S = 0 \cup S = 6 | p = 1/2) = (1/2)^3 = 1/8 = 12.5\%}$$

1.1.2 What is the power of the test for $H_a : p = 3/4$?

Power means supporting H_a assuming H_a is true, its associated probability is $1 - \beta = P(\text{support } H_a | H_a \text{ is true})$, where β is the probability of a type II error.

This is the same as the probability of our statistic falling in the rejection region assuming that $p = 3/4$, which expressed in terms of probability is the same as saying $(1 - \beta) = P(S = 0 \cup S = 6 | p = 3/4)$.

Because we already know that $P(S = 0 \cup S = 6) = p^3$, we only need to assume that $p = 3/4$ to get the desired probability:

$$\boxed{\text{Power} = (1 - \beta) = P(S = 0 \cup S = 6 | p = 3/4) = (3/4)^3 \approx 42.2\%}$$

1.2 Wrong Test, right data

1.2.1 What are the consequences of violating the metric scale assumption, if a paired t-test was used on this Likert data?

To accurately perform a paired t-test test on this data, one would need to assume that the response levels of the ordinal Likert variables are equally spaced apart. However, it's difficult to justify that the difference between "strongly agree" and "agree" is the same as the difference between "agree" and "neutral", and so on.

We also might struggle to draw any practical significance from the results. *A paired t-test is ultimately testing a difference in means between 2 observations on one set of individuals, but means do not make sense on Likert data.*

1.2.2 What would you propose to remedy this problem?

When considering a dependent test on this ordinal Likert data, a *signed test* may be a good option. The sign test does not require metric structure, but only records if the change in each paired case was positive or negative. The downside with the sign test is that we lose statistical power as we discard the magnitude of the changes in the paired data - a large sample size may be required to detect a statistically significant effect.

Other non-parametric dependent tests like the Wilcoxon signed-rank test may be more powerful, but still would require the same metric-scale assumptions that we are trying to avoid.

1.3 Statistical assumptions

1.3.1 World Happiness

The assumptions for running a two sample T-test are Metric Scale, IID data and Normality (or a large enough sample to count with CLT).

Metric Data

Upon evaluating the dataset happinees_WHR.csv, although the evaluation was on a scale of 1-10, the Cantril Ladder is an example of a Likert Scale meaning that we have a non-metric scale. In other words the values used in the dataset rely upon opinion (qualitative measurement) rather than a quantitative measurement. Therefore a two sample t-test fails on this assumption. View the query below to get an understanding of the Life Ladder variable.

```
library(readr)
happiness_WHR <- read_csv("./datasets/happiness_WHR.csv", show_col_types = FALSE)
summary(happiness_WHR$`Life Ladder`)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   2.375   4.971   5.768   5.678   6.428   7.889
```

IID Data Based on background knowledge of this study, each country is fairly independent of one another and therefore appears to fulfill this test assumption. Going one step further, there are no apparent violations of independence, some examples being clustering of data, in geographical regions, school cohorts, or families, strategical interaction, like competition among sellers or imitation of a species, or autocorrelation were one time period may affect the next. The data also appears to be identically distributed as will be shown in the plots below. This assumption for the data is meet, but as stated previously, a t-test would not be a wise hypothesis test to use for a non-metric statistical analysis.

No Major Deviations from Normality For further clarification, this assumption would not be met if the distribution was highly skewed for distributions when the sample size is larger than 30. In this problem, our data is slightly left skewed, although it does not meet the criteria for highly skewed data and therefore this assumption is meet.

Because the dataset doesn't categorize high and low GDP countries, we will categorize high GDP countries as those that have a GDP above the mean.

```
gdp_mean = mean(happiness_WHR$`Log GDP per capita`, na.rm = TRUE)
happiness_WHR = happiness_WHR %>%
  mutate(`GDP Category` = case_when(`Log GDP per capita` >= gdp_mean ~ "High GDP",
                                    `Log GDP per capita` < gdp_mean  ~ "Low GDP")) %>%
  filter(!is.na(`GDP Category`))
```

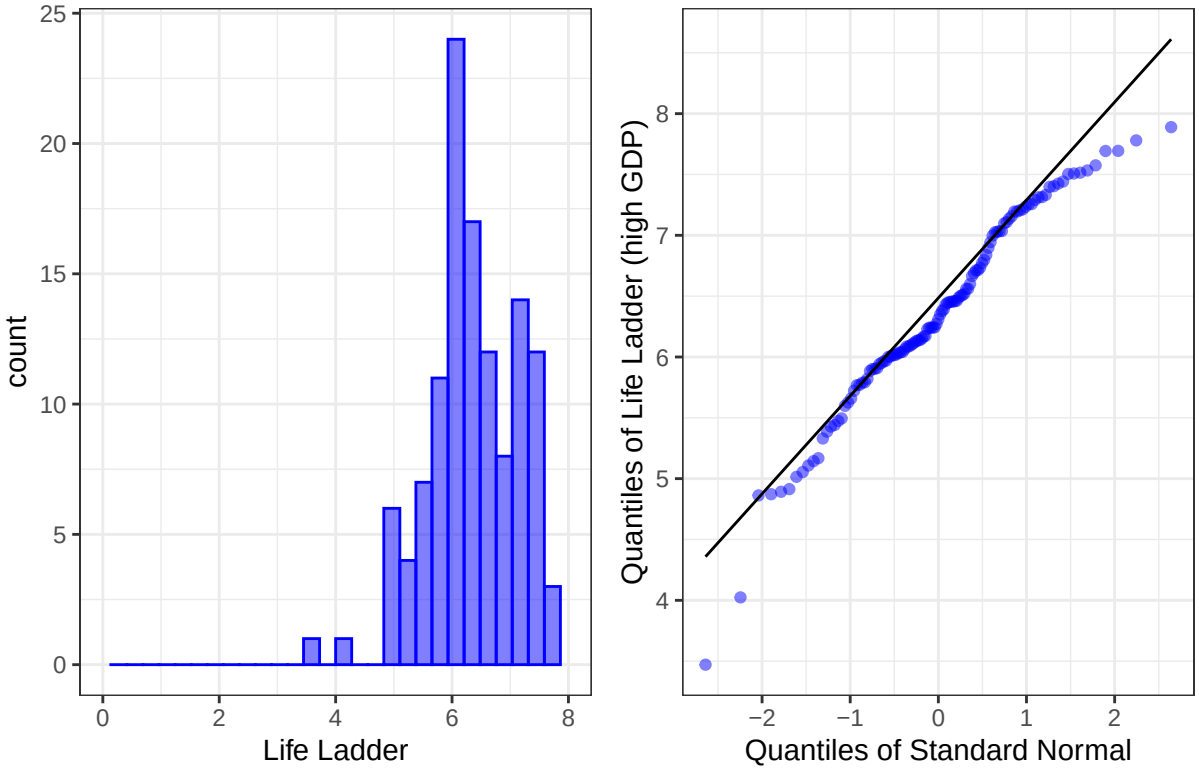
Below we can see the deviations from normality:

```
phigh_dist = happiness_WHR %>% filter(`GDP Category` == "High GDP") %>%
  ggplot() +
  aes(x=`Life Ladder`, fill=`GDP Category`, color=`GDP Category`) +
  geom_histogram(alpha=0.5, show.legend = FALSE) +
  scale_color_manual(values=c("blue")) + scale_fill_manual(values=c("blue")) +
  xlim(0, 8)

phigh_qq = happiness_WHR %>% filter(`GDP Category` == "High GDP") %>%
  ggplot() +
  aes(sample=`Life Ladder`) +
  stat_qq(color="blue", alpha=0.5) + stat_qq_line() +
```

```
  labs(
    x="Quantiles of Standard Normal",
    y="Quantiles of Life Ladder (high GDP)"
  )

(phigh_dist | phigh_qq) + plot_annotation(
  title="Distribution and QQ Plot for High GDP countries"
)
```

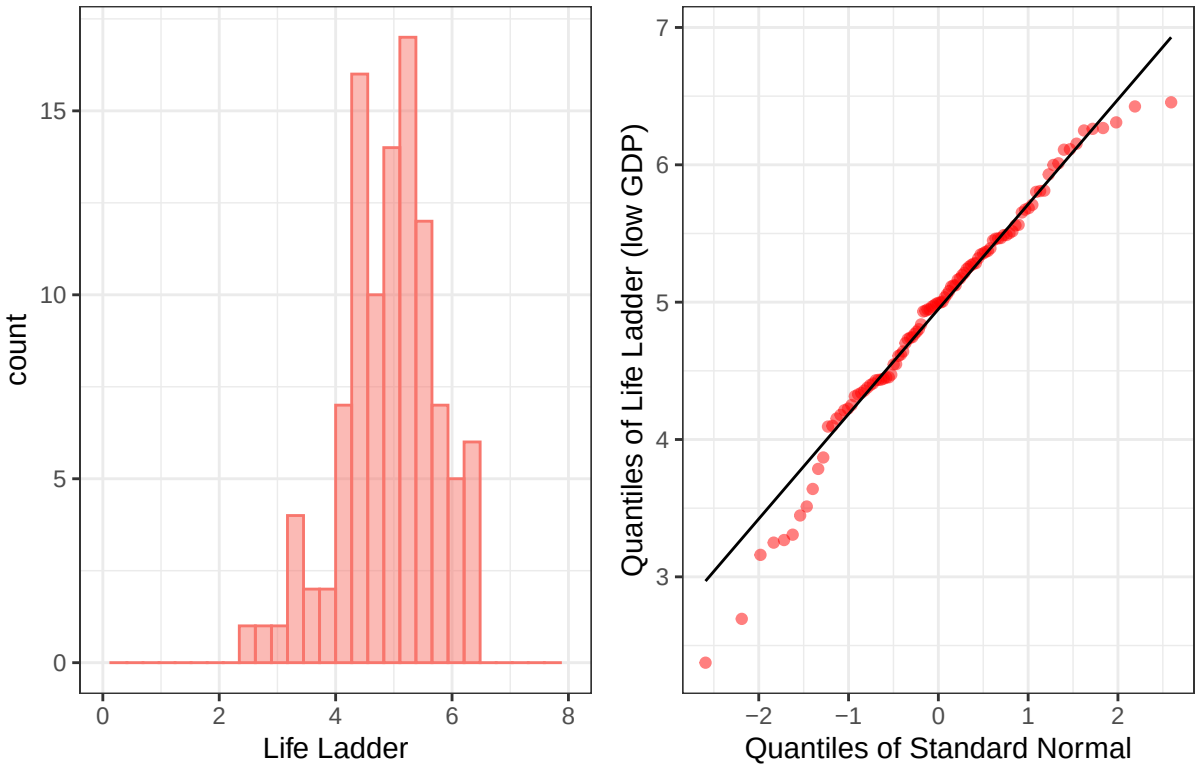



```
plow_dist = happiness_WHR %>% filter(`GDP Category` == "Low GDP") %>%
  ggplot() +
  aes(x=`Life Ladder`, fill=`GDP Category`, color=`GDP Category`) +
  geom_histogram(alpha=0.5, show.legend = FALSE) +
  xlim(0, 8)

plow_qq = happiness_WHR %>% filter(`GDP Category` == "Low GDP") %>%
  ggplot() +
  aes(sample=`Life Ladder`) +
  stat_qq(color="red", alpha=0.5) + stat_qq_line() +
  labs(
    x="Quantiles of Standard Normal",
    y="Quantiles of Life Ladder (low GDP)"
  )

(plow_dist | plow_qq) + plot_annotation(
  title="Distribution and QQ Plot for Low GDP countries"
)
```

Distribution and QQ Plot for Low GDP countries



We also ran the test, to see what it would yield:

```
t.test(happiness_WHR$`Life Ladder` ~ happiness_WHR$`GDP Category`)
```

```
##
##  Welch Two Sample t-test
##
## data:  happiness_WHR$`Life Ladder` by happiness_WHR$`GDP Category`
## t = 13.251, df = 218.47, p-value < 2.2e-16
## alternative hypothesis: true difference in means between group High GDP and group Low GDP is not equa
## 95 percent confidence interval:
##  1.218086 1.643727
## sample estimates:
## mean in group High GDP  mean in group Low GDP
##               6.355488                4.924581
```

Although we reject H_0 given this setup, a Two Sample T-Test would not be the appropriate hypothesis test based on the failed Metric Scale assumption. The other two assumptions are met.

1.3.2 Legislators

The Wilcoxon Rank-Sum Test (Hypothesis of Comparisons) tests if the probability of drawing a rank from certain group is the same as drawing a rank from another group ($H_0 : P(X > Y) = P(Y > X)$) and it has 2 main assumptions: Ordinal scale and IID data

Ordinal Scale

Based on background knowledge, ordinal scale measures data of categorical nature where ordered categories and the distances between the categories are not known. Since in this problem we are measuring age given

by the birthday of each congressman, and age would be considered a metric variable, this test would fail on this assumption and not be a viable option for statistical analysis. View a summary of the data below.

```
library(readr)
legislators_current <- read_csv("./datasets/legislators-current.csv", show_col_types = FALSE)
summary(legislators_current)
```

```
##   last_name          first_name        middle_name           suffix
##  Length:538         Length:538         Length:538         Length:538
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##    nickname          full_name            birthday             gender
##  Length:538         Length:538         Min.   :1933-06-09   Length:538
##  Class :character   Class :character   1st Qu.:1953-03-30   Class :character
##  Mode  :character   Mode  :character   Median :1961-03-05   Mode  :character
##                                        Mean   :1961-11-29
##                                        3rd Qu.:1970-08-14
##                                        Max.   :1995-08-01
##
##      type               state              district       senate_class
##  Length:538         Length:538         Min.   : 0.000   Min.   :1.00
##  Class :character   Class :character   1st Qu.: 3.000   1st Qu.:1.00
##  Mode  :character   Mode  :character   Median : 6.000   Median :2.00
##                                        Mean   : 9.984   Mean   :2.01
##                                        3rd Qu.:13.000   3rd Qu.:3.00
##                                        Max.   :53.000   Max.   :3.00
##                                        NA's   :100      NA's   :438
##     party                url              address             phone
##  Length:538         Length:538         Length:538         Length:538
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##  contact_form         rss_url            twitter            facebook
##  Length:538         Length:538         Length:538         Length:538
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##    youtube           youtube_id         bioguide_id         thomas_id
##  Length:538         Length:538         Length:538         Length:538
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
```

```
##  opensecrets_id        lis_id            fec_ids             cspan_id
##  Length:538         Length:538         Length:538         Min.   :    260
##  Class :character   Class :character   Class :character   1st Qu.:  45591
##  Mode  :character   Mode  :character   Mode  :character   Median :  79718
##                                                           Mean   : 543374
##                                                           3rd Qu.:1003305
##                                                           Max.   :9275683
##                                                           NA's   :156
##   govtrack_id      votesmart_id    ballotpedia_id     washington_post_id
##  Min.   :300018   Min.   :   119   Length:538         Mode:logical
##  1st Qu.:412199   1st Qu.: 22411   Class :character   NA's:538
##  Median :412570   Median : 52964   Mode  :character
##  Mean   :412042   Mean   : 75411
##  3rd Qu.:412772   3rd Qu.:133024
##  Max.   :456862   Max.   :188334
##                   NA's   :62
##     icpsr_id     wikipedia_id
##  Min.   :14066   Length:538
##  1st Qu.:21106   Class :character
##  Median :21564   Mode  :character
##  Mean   :24264
##  3rd Qu.:21972
##  Max.   :94659
##  NA's   :77
```

IID Data For this assumption, each X_i has to be drawn from the same distribution, each Y_i has to be drawn from the same distribution, and all X_i and Y_i are independent. There are no apparent violations of independence, such as clustering of data, in geographical regions, school cohorts, or families, strategical interaction, like competition among sellers or imitation of a species, or autocorrelation were one time period may affect the next. The data also appears to be identically distributed as will be shown in the plots below. This assumption for the data is meet, but as stated previously, the Hypothesis of Comparison version of the Wilcoxon Rank-Sum would not be a wise test to use for a metric statistical analysis.

For the analysis of this dataset, a simple extraction of the each congressman's birth year subtracted from the current year yields their age.

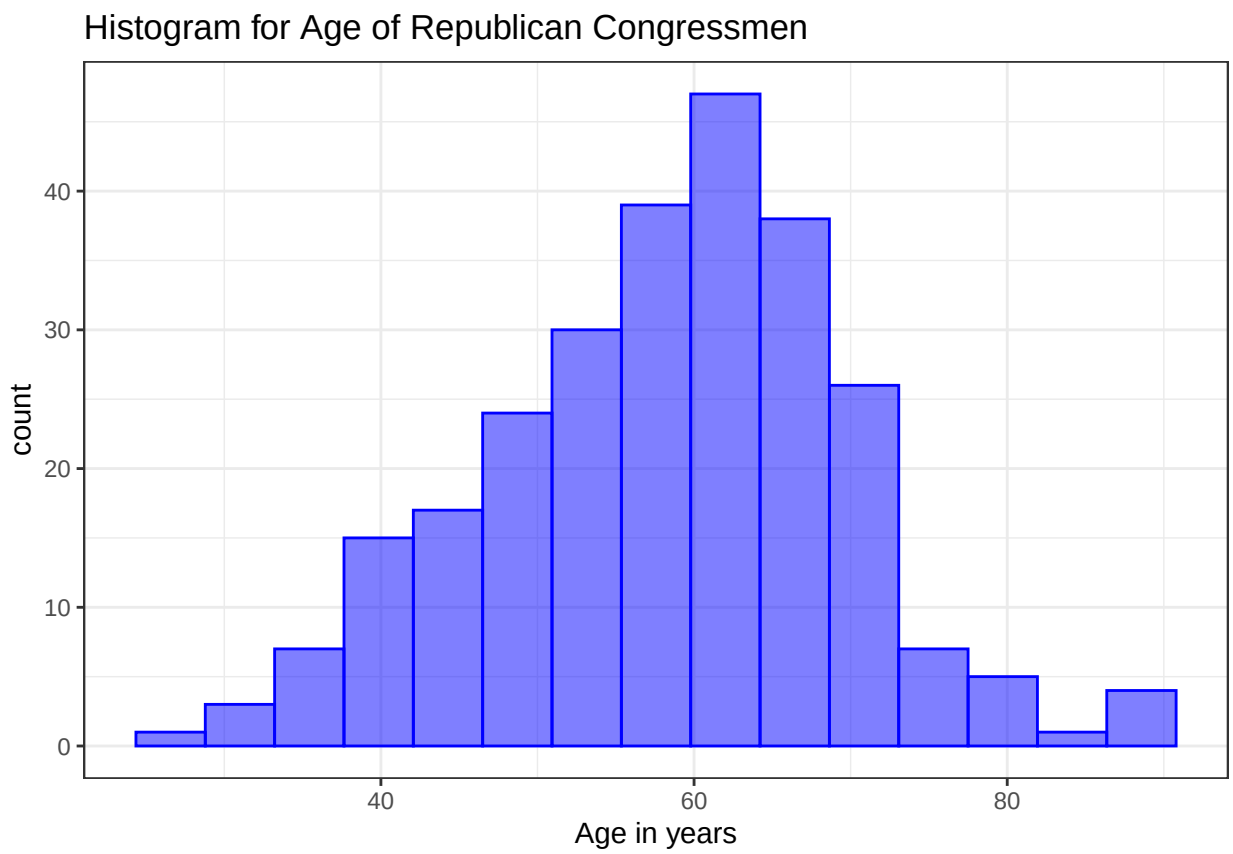
```
a <- as.POSIXct(legislators_current$birthday, format = "%Y-%m-%d")
year <- strtoi(format(a, format = "%Y"))
legislators_current$year = year
legislators_current$age <- 2021 - legislators_current$year
```

Next, subsetting the data by Republicans and Democrats in rep and dem respectively.

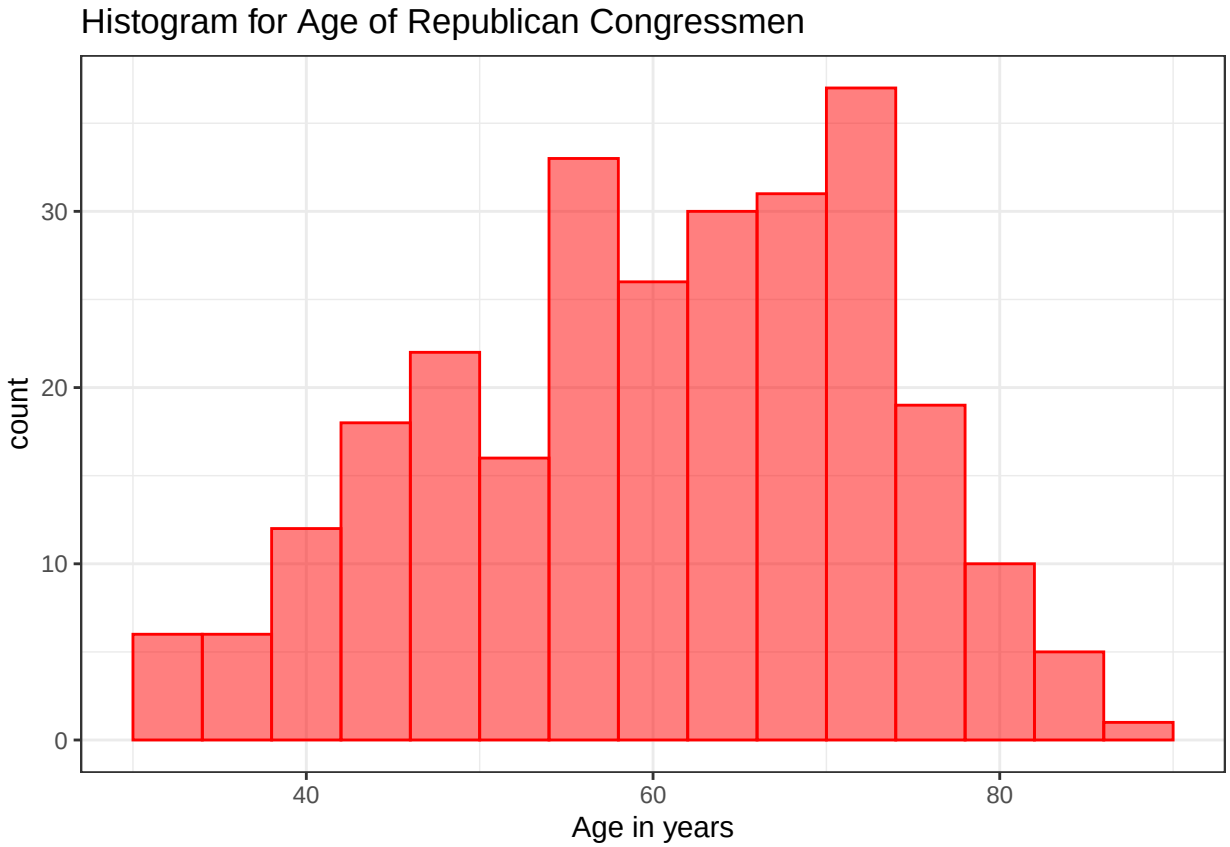
```
legislators_current$party[legislators_current$party == 'Independent'] = NA
rep <- subset(legislators_current, subset = legislators_current$party == "Republican")
dem <- subset(legislators_current, subset = legislators_current$party == "Democrat")
```

And finally, plotting the resulting of ages for Republicans and Democrats.

```
ggplot() +
  aes(x=rep$age) +
  geom_histogram(alpha=0.5, fill='blue', color='blue', bins=15) +
  labs(
    title="Histogram for Age of Republican Congressmen",
    x="Age in years"
  )
```




```
ggplot() +
  aes(x=dem$age) +
  geom_histogram(alpha=0.5, fill='red', color='red', bins=15) +
  labs(
    title="Histogram for Age of Republican Congressmen",
    x="Age in years"
  )
```



If we run the Wilcoxon Rank-Sum Test, we can see that under these conditions, we would reject H_0

```
wilcox.test(rep$age, dem$age)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  rep$age and dem$age
## W = 31022, p-value = 0.006447
## alternative hypothesis: true location shift is not equal to 0
```

In conclusion, the Wilcoxon Rank-Sum Test (Hypothesis of Comparisons) would not be the appropriate hypothesis test based on the failed Ordinal Scale assumption. The other assumptions are met.

1.3.3 Wine and Health

The assumptions for Wilcoxon Signed-Rank Test include Metric scale, IID Data and the distribution of the difference $(X - Y)$ is symmetric around the same mean

Metric Data For clarification, the X and Y measured here have to be measured on the same scale since we are using a paired test. For this question, both the liver and heart deaths are on the same scale of 100,000 deaths. Therefore this assumption is met. This will be shown in the dataset below.

```
library(wooldridge)
summary(wine)
```

```
##    country             alcohol          deaths         heart
##  Length:21          Min.   :0.600   Min.   : 680   Min.   : 36.0
##  Class :character   1st Qu.:1.200   1st Qu.: 751   1st Qu.:131.0
```

```
##  Mode  :character   Median :1.900   Median : 806   Median :191.0
##                     Mean   :2.838   Mean   : 830   Mean   :183.3
##                     3rd Qu.:2.900   3rd Qu.: 916   3rd Qu.:220.0
##                     Max.   :9.100   Max.   :1000   Max.   :300.0
##      liver
##  Min.   : 6.50
##  1st Qu.:11.20
##  Median :19.00
##  Mean   :21.03
##  3rd Qu.:23.90
##  Max.   :45.60
```

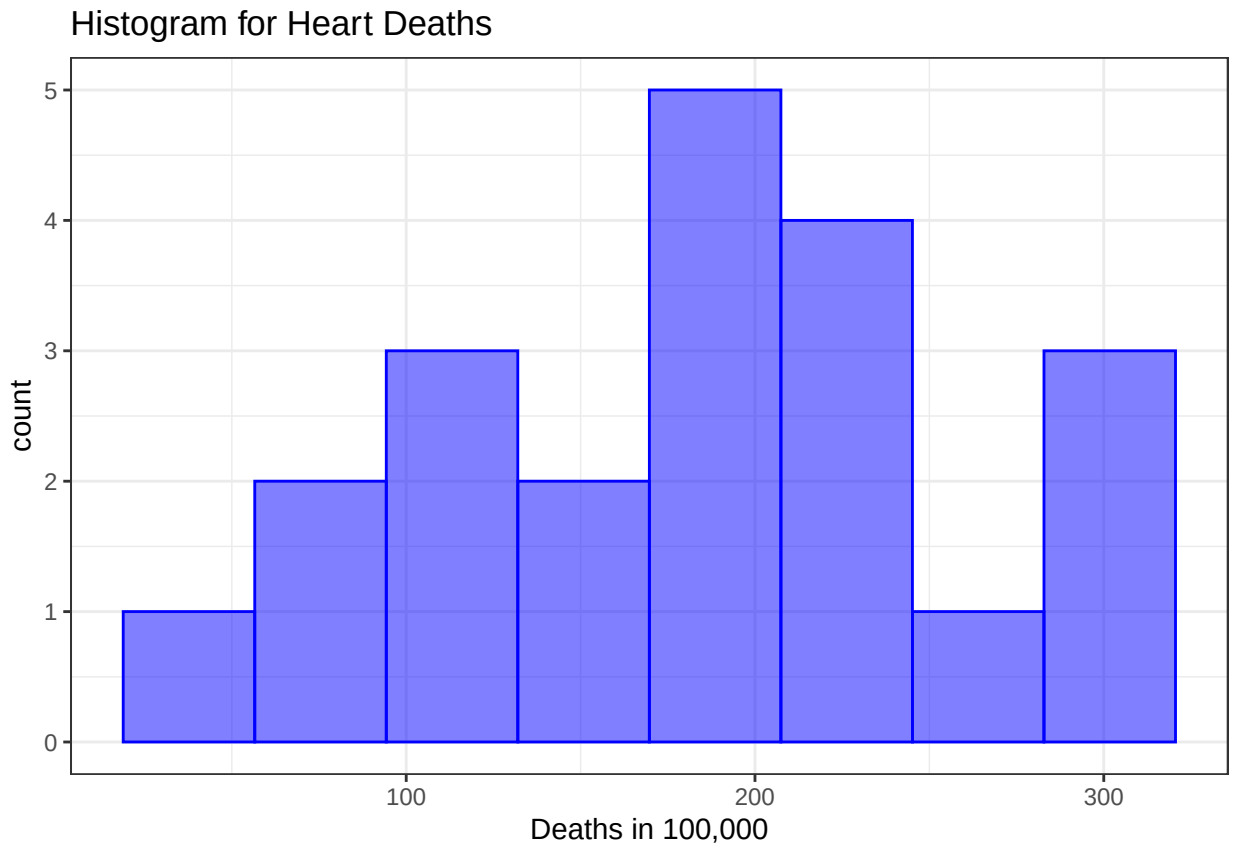
IID Data For further clarification of this assumption, each pair (X_i, Y_i) has to be drawn from the same distribution independently of all other pairs. Even though this problem has a small dataset, it appears to be independent due to each observation being a different country and the data also appears to be identically distributed as will be shown in the plots below. In this case, There are no apparent violates of independents, some examples being clustering of data, in geographical regions, school cohorts, or families, strategical interaction, like competition among sellers or imitation of a species, or autocorrelation were one time period may affect the next. This assumption for the data is met and thus far, the Wilcoxon Ranked-Summed Test appears to be a viable test.

For the data visualization, two histograms of heart deaths and liver deaths were made in the plots below.

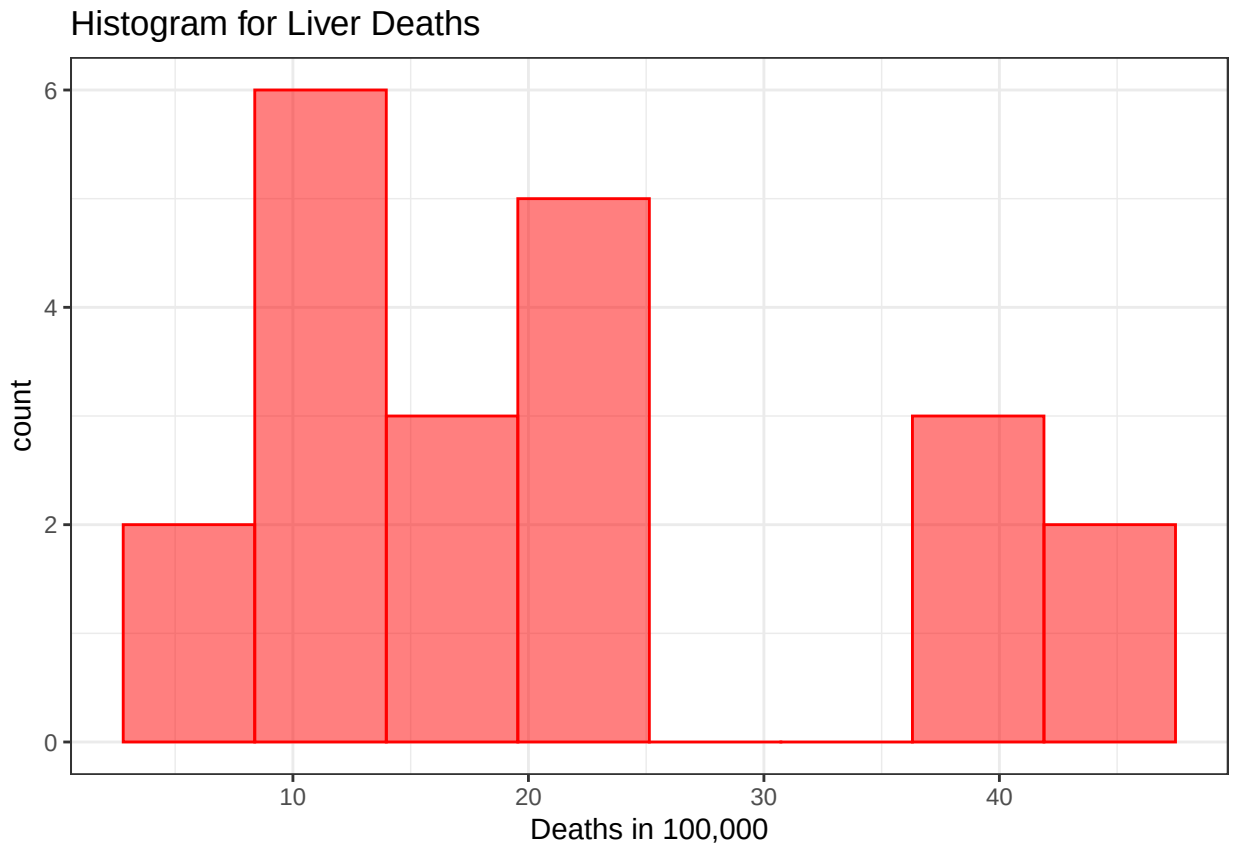
```
liver_m <- mean(wine$liver)
heart_m <- mean(wine$heart)
wine$difference <- wine$heart - wine$liver
(heart_m - liver_m)
```

```
## [1] 162.2524
```

```
ggplot() +
  aes(x=wine$heart) +
  geom_histogram(alpha=0.5, color='blue', fill="blue", bins=8) +
  labs(
    title="Histogram for Heart Deaths",
    x="Deaths in 100,000"
  )
```

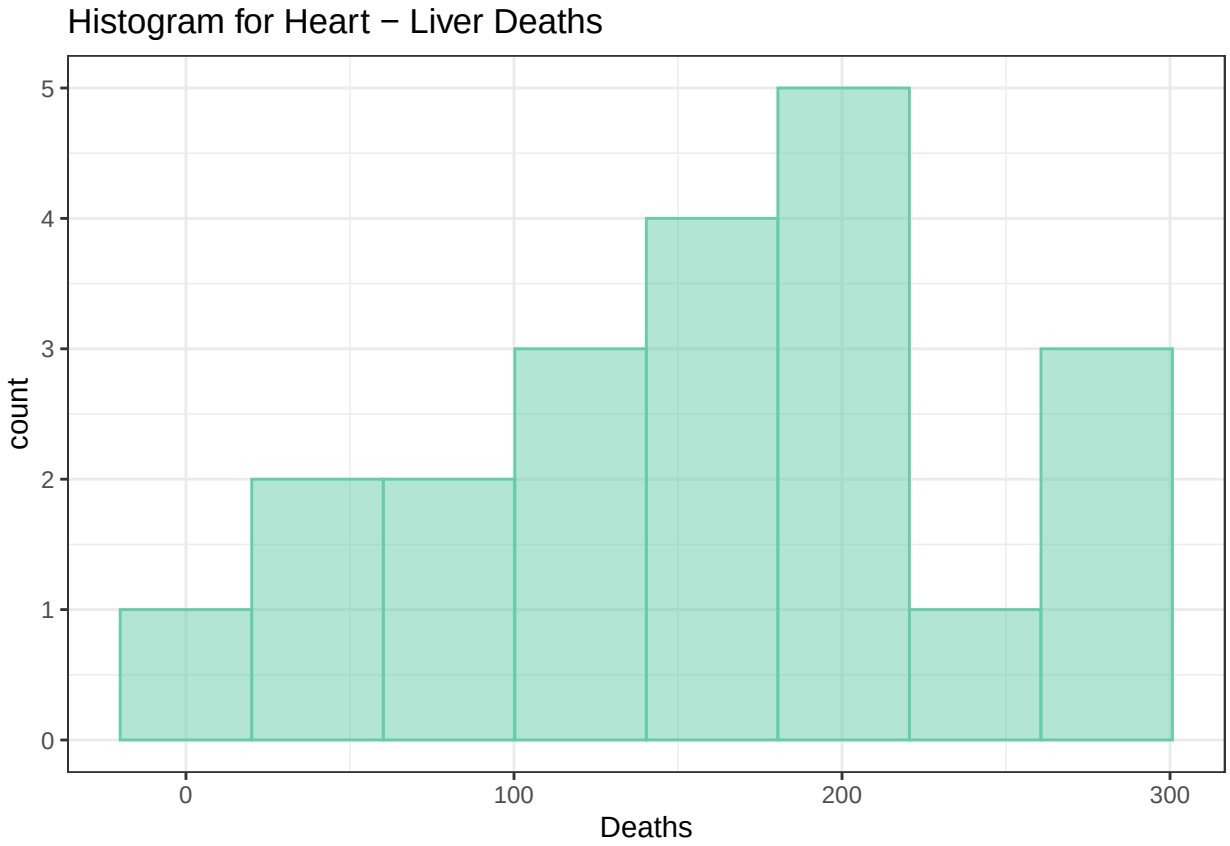



```
ggplot() +
  aes(x=wine$liver) +
  geom_histogram(alpha=0.5, color='red', fill="red", bins=8) +
  labs(
    title="Histogram for Liver Deaths",
    x="Deaths in 100,000"
  )
```



The Distribution of the difference $(X - Y)$ is the Same Around Some Mean Using the data visualization below, we have taken the difference of Heart and Liver deaths over the 21 different observations and take the difference of their means (162) as the point for the data to be symmetrical around. The data below shows that even with this small dataset it appears that the data meets this assumption being symmetrical around the mean of 162.

```
ggplot() +
  aes(x=wine$difference) +
  geom_histogram(alpha=0.5, color='aquamarine3', fill="aquamarine3", bins=8) +
  labs(
    title="Histogram for Heart - Liver Deaths",
    x="Deaths"
  )
```



The test was ran below to see what the results might be. Due to the low p-value, we would be inclined to reject the null hypothesis.

```
wilcox.test(wine$heart, wine$liver, paired=TRUE)
```

```
##
##  Wilcoxon signed rank exact test
##
## data:  wine$heart and wine$liver
## V = 231, p-value = 9.537e-07
## alternative hypothesis: true location shift is not equal to 0
```

In conclusion, all of the assumptions have been met for the Wilcoxon Signed-Rank Test and it appears to be a viable hypothesis test to use for this dataset.

1.3.4 Attitudes Towards Religion

The paired T-test assumptions are 3: Metric Data, IID Data and no major deviations from normality.

The H_0 that this test is trying to falsify is that the expectation of X equals the expectation of Y, when X and Y are measurements from the same individual.

Metric Scale For further clarification, the t-test is not valid for variables which only have an ordinal structure. In this case, the feeling thermometer used to determines one's feelings towards Catholics or Protestants would fall under a Likert Scale type of variable and also be considered of ordinal structure. Therefore, the test would fail under this assumption. A summary of the given data is shown below as supporting evidence.

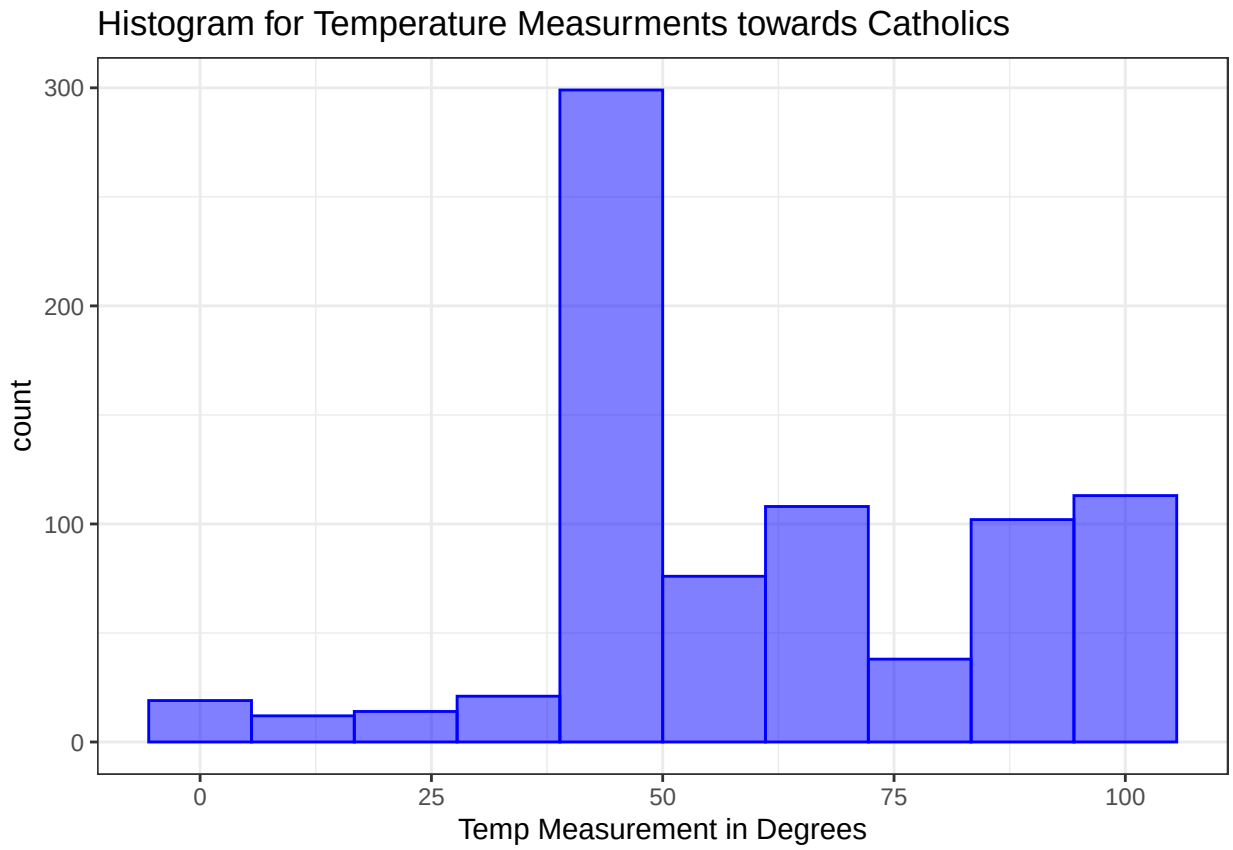
```
GSS <- read_csv("./datasets/GSS_religion.csv", show_col_types = FALSE)
summary(GSS)
```

```
##       ...1            year            id            prottemp
##  Min.   :  1.0   Min.   :2004   Min.   :   4.0   Min.   :  0.00
##  1st Qu.:201.2   1st Qu.:2004   1st Qu.: 728.8   1st Qu.: 50.00
##  Median :401.5   Median :2004   Median :1373.5   Median : 60.00
##  Mean   :401.5   Mean   :2004   Mean   :1381.9   Mean   : 65.56
##  3rd Qu.:601.8   3rd Qu.:2004   3rd Qu.:2053.5   3rd Qu.: 85.00
##  Max.   :802.0   Max.   :2004   Max.   :2808.0   Max.   :100.00
##     cathtemp
##  Min.   :  0.00
##  1st Qu.: 50.00
##  Median : 60.00
##  Mean   : 63.16
##  3rd Qu.: 85.00
##  Max.   :100.00
```

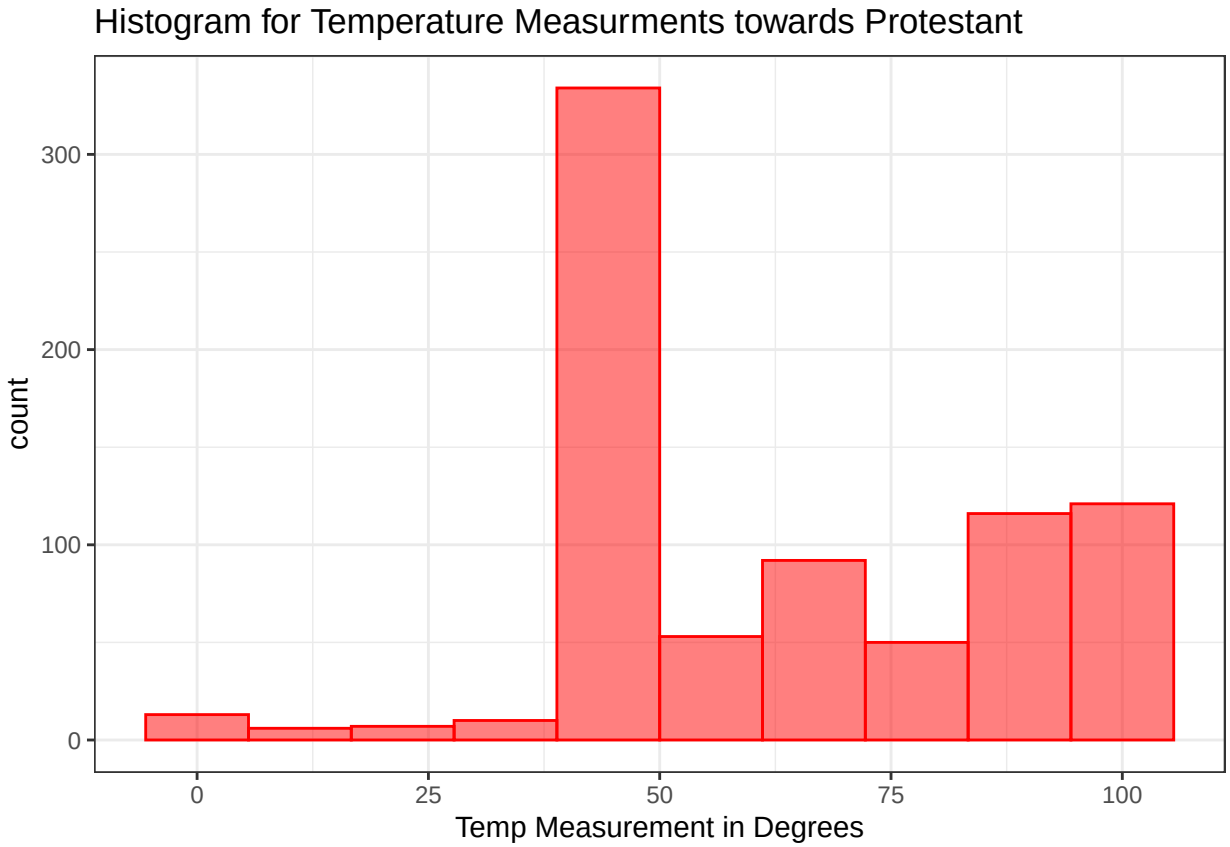
IID Data Making use of background knowledge, each pair of measurements (X_i, Y_i) is drawn from the same distribution, independently of all other pairs. This data appears to have paring, independence and identical distribution are not violated. There are no apparent violates of independents, some examples being clustering of data, in geographical regions, school cohorts, or families, strategical interaction, like competition among sellers or imitation of a species, or autocorrelation were one time period may affect the next. The data also appears to be identically distributed as will be shown in the plots below. This assumption for the data is meet, but as stated previously, the Paired t-Test would not be a wise test to use for an ordinal scale statistical analysis.

No Major Deviations from Noramlity, considering the sample size For clarification, the t-test is invalid for highly skewed distributions with large sample sizes. This does not appear to be the case as shown in the graphs below. This assumption for the data is meet, but as stated previously, the Paired t-Test would not be a wise test to use for an ordinal scale statistical analysis. See the histograms below for as supporting evidence.

```
ggplot() +
  aes(x=GSS$cathtemp) +
  geom_histogram(color="blue", fill="blue", alpha=0.5, bins=10) +
  labs(
    title="Histogram for Temperature Measurments towards Catholics",
    x="Temp Measurement in Degrees"
  )
```




```
ggplot() +
  aes(x=GSS$prottemp) +
  geom_histogram(color="red", fill="red", alpha=0.5, bins=10) +
  labs(
    title="Histogram for Temperature Measurments towards Protestant",
    x="Temp Measurement in Degrees"
  )
```



A Paired T-Test was run out of curiosity for the results:

```
t.test(GSS$prottemp, GSS$cathtemp, paired=TRUE)
```

```
##
##  Paired t-test
##
## data:  GSS$prottemp and GSS$cathtemp
## t = 2.9249, df = 801, p-value = 0.003543
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.7902444 4.0152419
## sample estimates:
## mean of the differences
##                2.402743
```

Although we reject H_0, we conclude that the Paired T-Test would not be the appropriate hypothesis test based on the failed Metric Scale assumption. The other assumptions are met.

Conclusion

In conclusion, through our analysis of the different scenarios where hypothesis tests could be run, the World Happiness statistical test would not be best evaluated with a Two-Sample T-Test, the Legislators problem would not be best evaluated with a hypothesis of comparisons version of the Wilcoxon Rank-Sum Test, the Wine and Health problem does fit all the assumptions for a Wilcoxon Signed-Rank Test and finally the Attitudes towards Religions problem does not fit the assumptions for a Paired T-Test.