

Final Project: An Evaluation of the World Happiness Report

Anna MacFarlane, Edgar Salas, Jerry Fu, Taalin RaoShah

11/4/2020

Introduction

CodeStone chose to explore the data from the World Happiness Reports of 2018 and 2019, published by the Sustainable Development Network. According to the WHR, this data is derived from the Gallup World Poll, systematic telephone surveys and in person interviews in over 160 countries whose surveys claim to represent 80%+ of the population (2014). The calls are made via random phone number generation and randomly selecting households (GWP, 2014). The data sets from each year include 312 cases, each representing the data for a particular country. The original variables in the data set included the overall rank according to happiness score, country name, happiness score, and then 6 variables that were used to calculate the happiness score in the WHR's analysis. These are GDP per capita, social support, healthy life expectancy, freedom to make life choices, generosity, and perceptions of corruption, all of which are expanded upon in the methodology section of this report. Additionally, we added variables for year as well as region to further our analysis.

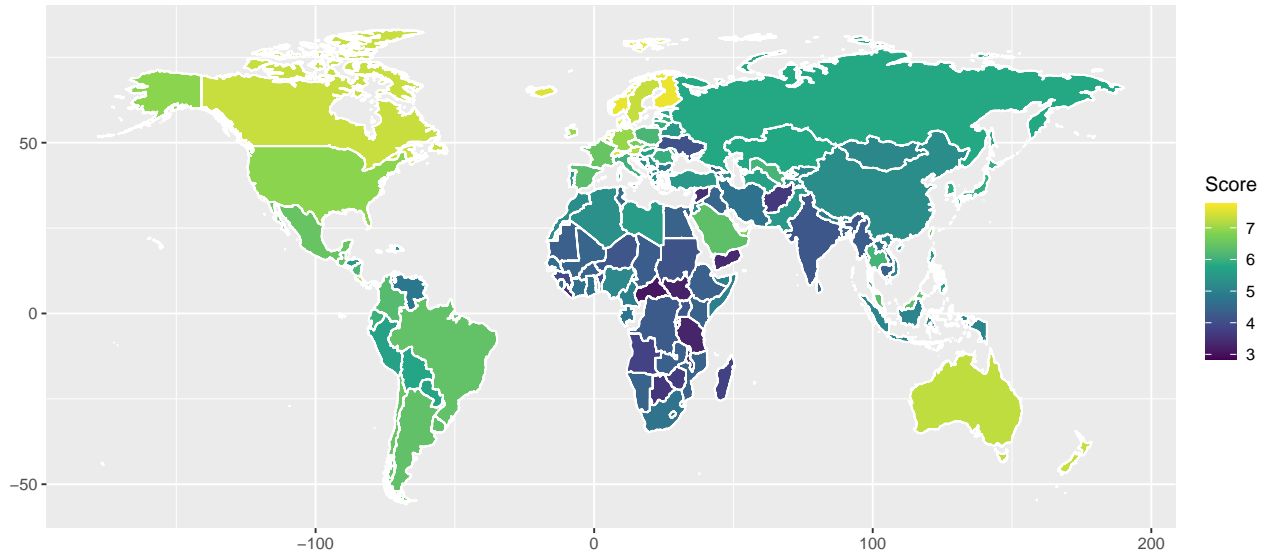
Generally, we want to explore the following question: what factors contribute to the differences in happiness scores between countries in North America and Australia versus Western Europe? Both regions are relatively comparable in terms of financial and political structures, but Western Europe contains the Nordic countries which consistently score highest in the World Happiness reports. With this, questions are raised as to what contributes to such happiness in this region of the world.

Considering the concerning state of our world in 2020, including the worsening effects of climate change, threats to democracy, and much more, we found it topical and insightful to evaluate what contributes to happiness within nations and between regions across the globe. The first World Happiness Report, published in 2012, presents the report as a means of grappling with the countless contradictions that exist in modern society such as the balance between pursuing economic success versus protecting the environment or the tradeoffs between personal profit and community trust (Helliwell et al., 2012). Eight years later, these paradoxes persist, and the potential solutions are closely linked to definitions of morality, heightening their controversy. Considering the continued debate over such questions, we believe there are grounds for further investigation into trends of happiness over time and the factors that contribute to it.

```
## Rows: 312
## Columns: 11
## $ Overall_rank      <dbl> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12,...
## $ Country           <chr> "Finland", "Norway", "Denmark", "Icela...
## $ Score             <dbl> 7.632, 7.594, 7.555, 7.495, 7.487, 7.4...
## $ GDP_per_capita    <dbl> 1.305, 1.456, 1.351, 1.343, 1.420, 1.3...
## $ Social_support     <dbl> 1.592, 1.582, 1.590, 1.644, 1.549, 1.4...
## $ Healthy_life_expectancy <dbl> 0.874, 0.861, 0.868, 0.914, 0.927, 0.8...
## $ Freedom_to_make_life_choices <dbl> 0.681, 0.686, 0.683, 0.677, 0.660, 0.6...
## $ Generosity         <dbl> 0.202, 0.286, 0.284, 0.353, 0.256, 0.3...
## $ Perceptions_of_corruption <dbl> 0.393, 0.340, 0.408, 0.138, 0.357, 0.2...
## $ year              <chr> "2018", "2018", "2018", "2018", "2018"...
## $ Region            <chr> "Western Europe", "Western Europe", "W...
```

Is there a relationship between region of the world and happiness score?

World Map
colored by Happiness Score



<https://www.datanovia.com/en/blog/how-to-create-a-map-using-ggplot2/>

H_0 : There is no association between region of the world and happiness score. H_a : There is an association between region of the world and happiness score.

```
##  
## Pearson's Chi-squared test with simulated p-value (based on 2000  
## replicates)  
##  
## data: worldhappiness$Score and worldhappiness$Region  
## X-squared = 2562.2, df = NA, p-value = 0.1099
```

With a simulated p-value of 0.0004998 which is less than our $\alpha = 0.05$, we have sufficient evidence to reject the null hypothesis in favor of the alternative hypothesis and conclude that there is an association between region of the world and happiness score.

Summary Statistics

```
## WE_mean_happiness median_score      SD  
## 1          6.86375          7.003 0.6939077  
  
## NA_mean_happiness median_score      SD  
## 1          7.189375          7.275 0.1881557
```

Each of the simulation-based hypothesis tests used to evaluate the variables contributing to overall happiness score will be evaluated at the $\alpha = 0.05$ level.

Generally, our null hypothesis states that the mean score for a particular variable in Western Europe is less than or equal to the mean score for that same variable in North America, Australia, and New Zealand. Our alternative hypothesis therefore is that the mean score for a particular variable in Western Europe is greater than that of the mean score for that variable in North America, Australia, and New Zealand.

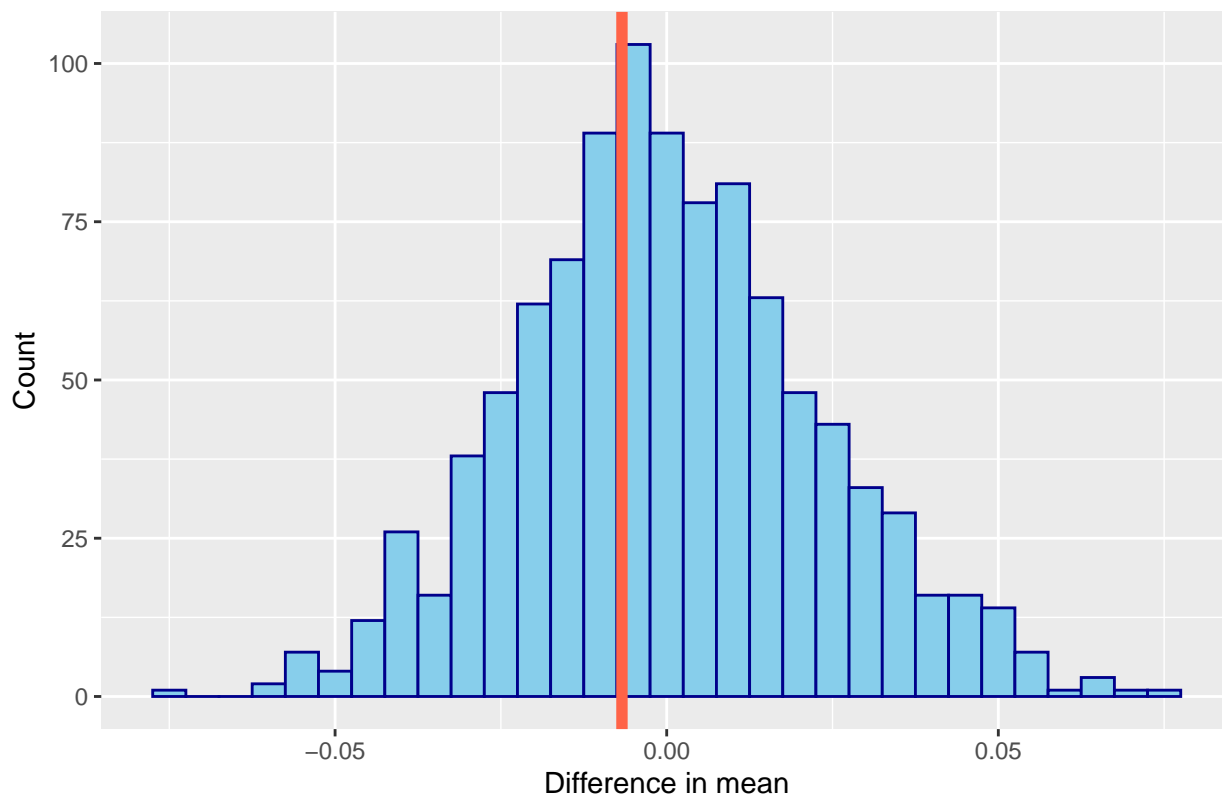
$H_0: \mu_{WE} \leq \mu_{NA,ANZ}$

$H_A: \mu_{WE} > \mu_{NA,ANZ}$

GDP per capita

```
## WE_GDP
## 1 1.34385
## NA_ANZ_GDP
## 1 1.351125
## p_val
## 1 0.366
```

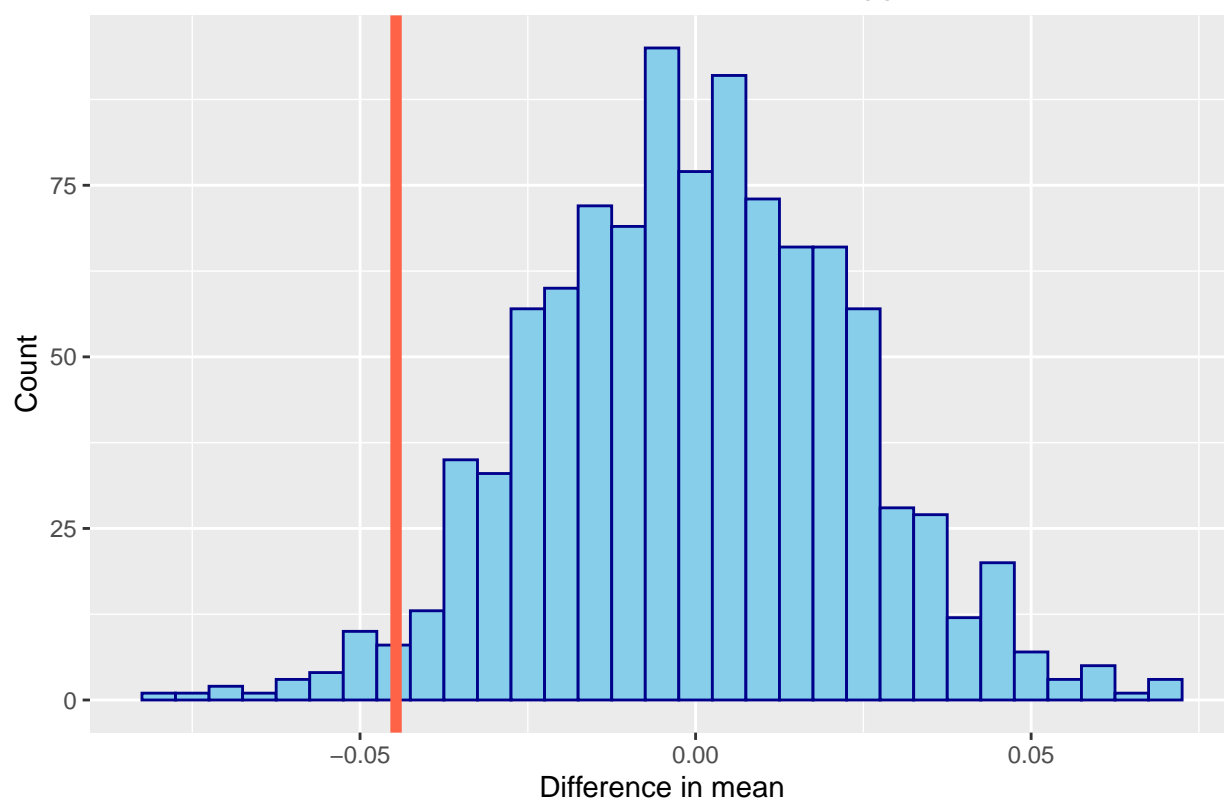
Simulated null distribution of difference in GDP per capita



Social Support

```
## WE_Soc_Support
## 1 1.486825
## NA_ANZ_Soc_Support
## 1 1.5305
## p_val
## 1 0.03
```

Simulated null distribution of difference in social support



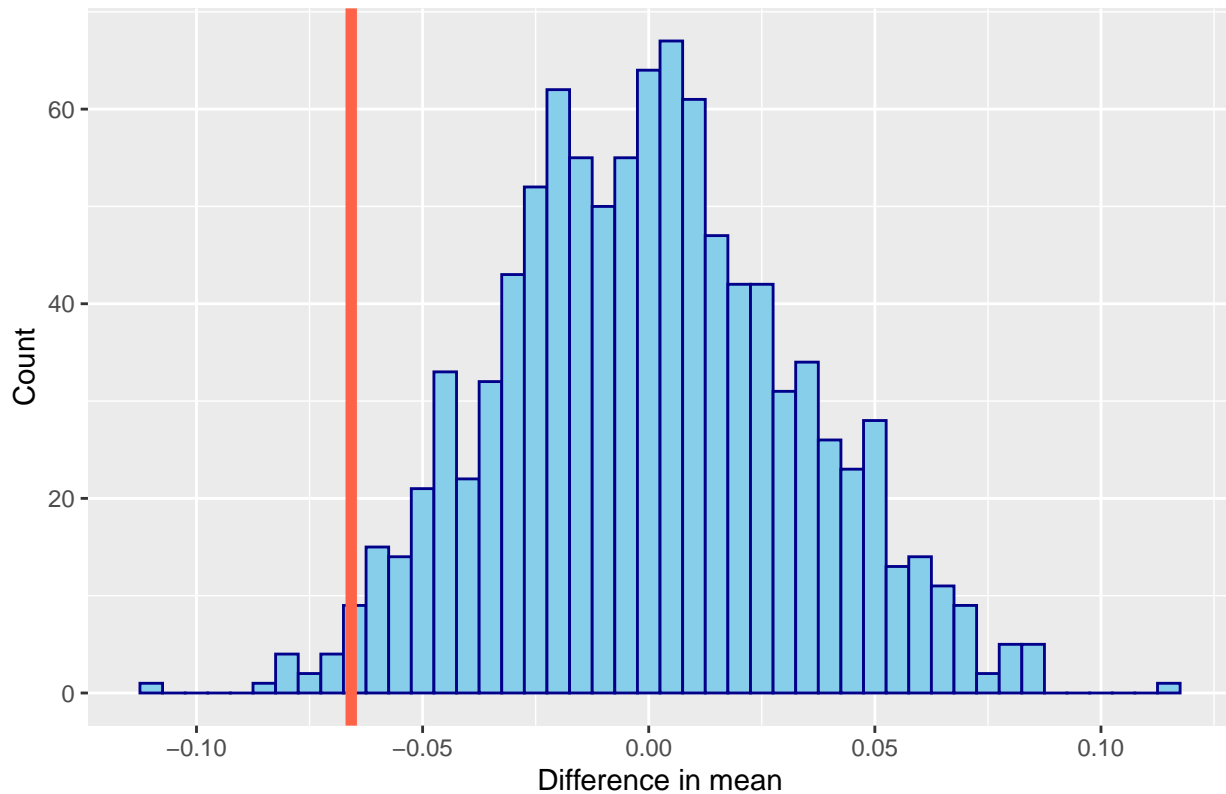
Freedom to make life choices

```
## WE_Life_Choices
## 1 0.52105

## NA_ANZ_Life_Choices
## 1 0.587

## p_val
## 1 0.024
```

Simulated null distribution of difference in freedom to make life choices



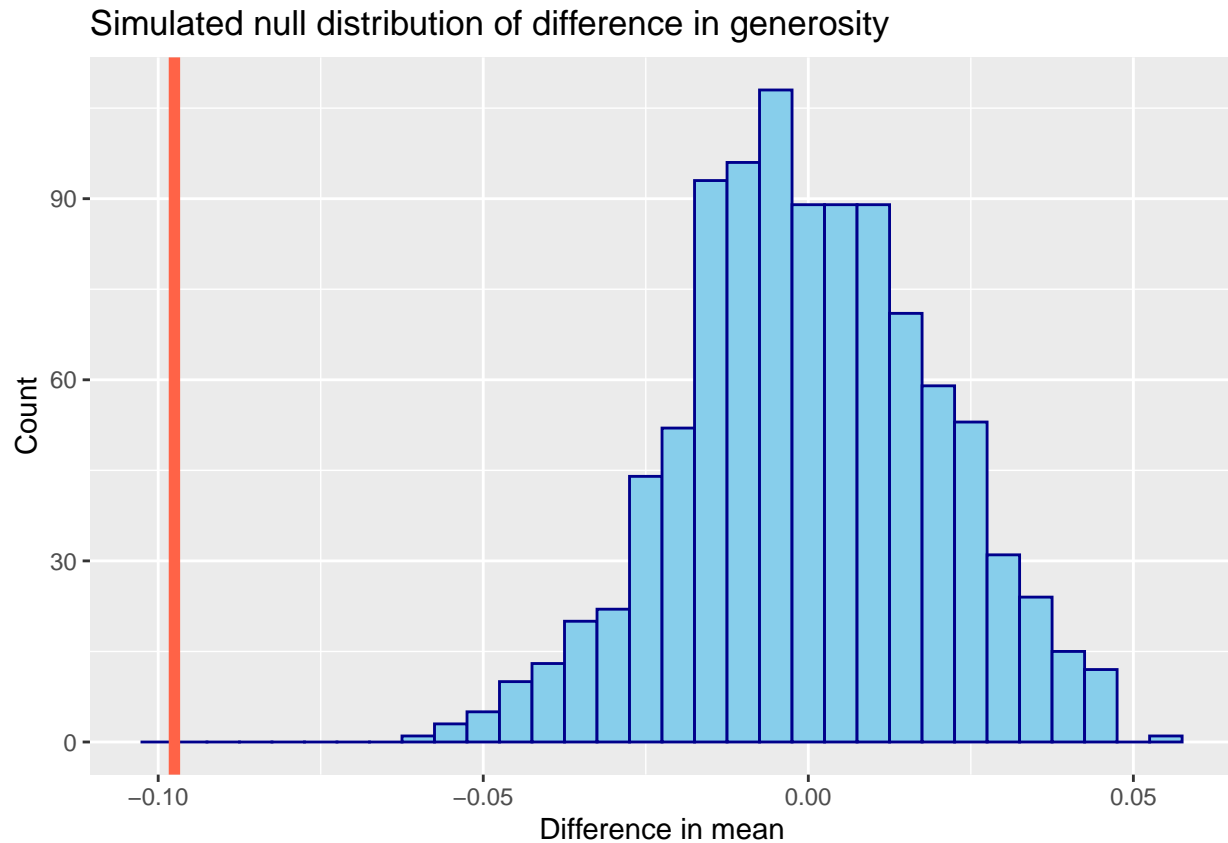
The calculated p-value of 0.024 is significant at the $\alpha = 0.05$ level, indicating that we can reject the null hypothesis. Therefore, there is statistically significant evidence to suggest that the mean freedom to make life choices score in Western Europe is greater than that of North America, Australia, and New Zealand.

Generosity

```
## WE_Gen
## 1 0.22245

## NA_ANZ_Gen
## 1 0.320625

## p_val
## 1 0
```



The calculated p-value of 0 is significant at the $\alpha = 0.05$ level, indicating that we can reject the null hypothesis. Therefore, there is statistically significant evidence to suggest that the mean generosity score in Western Europe is greater than that of North America, Australia, and New Zealand.

Life Expectancy

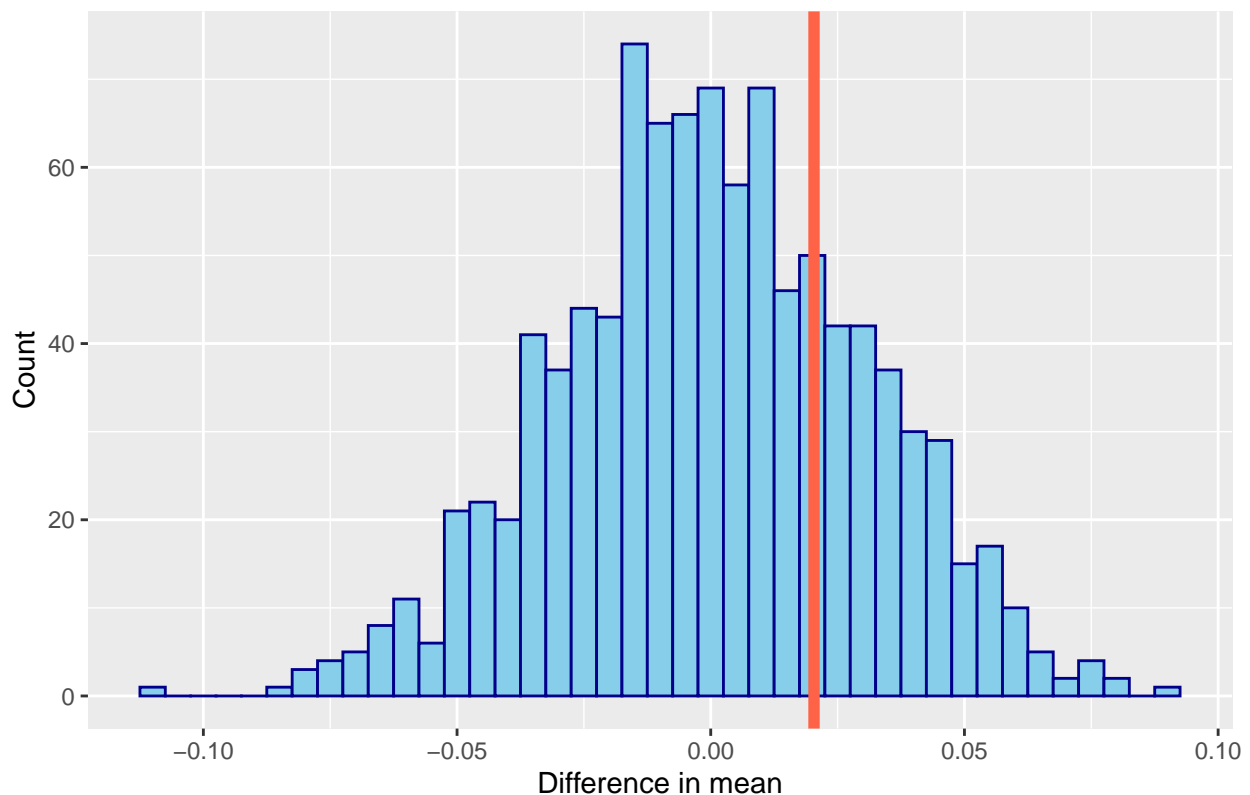
H_o : The mean life expectancy in North America is less than or equal to that of Western Europe.

H_a : The mean life expectancy in North America is greater than that of Western Europe.

The life expectancies for the two regions are quite similar and a p-value of 0.264 means we fail to reject the null hypothesis. This supports the claim that there is no significant difference in life expectancy between the two regions.

```
## WE_life_expectancy
## 1 0.954825
## RestofW_life_expectancy
## 1 0.954825
## p_val
## 1 0.262
```

Simulated null distribution of difference in life expectancy



Perceptions of Corruption

H_o : The mean perceptions of corruption in North America is less than or equal to that of Western Europe.

H_a : The mean perceptions of corruption in North America is greater than that of Western Europe.

$\alpha = 0.05$

From this test, we find a P-value of 0.0535. Since our p-value is not less than our significance level of $\alpha = 0.05$, we fail to reject the null hypothesis. There is not enough evidence to suggest that the mean perceptions of corruption in North America is greater than that of Western Europe.

```
#Set seed
set.seed(5318008)

#Summary Stats
worldhappiness %>%
  filter(Region == "Western Europe") %>%
  summarize(WE_POC = mean(Perceptions_of_corruption))
```

```
##      WE_POC
## 1 0.217675
```

```
worldhappiness %>%
  filter(Region == "North America and ANZ") %>%
  summarize(NA_POC = mean(Perceptions_of_corruption))
```

```
##      NA_POC
## 1 0.277625
```

```
bootstrap <- diff_means(variable = "Perceptions_of_corruption", 2000)

bootstrap$p_val

##      p_val
## 1 0.0535

#Graph
ggplot(bootstrap$null_dist, aes(x = centered_diffs)) +
  geom_histogram(binwidth = 0.01, fill = "skyblue", color = "darkblue") +
  geom_vline(xintercept = bootstrap$obs_diff, color = "tomato", lwd = 2) +
  labs(title = "Simulated Null Distribution of Difference in mean Perceptions of Corruption",
       x = "Difference in Mean",
       y = "Count")
```

Simulated Null Distribution of Difference in mean Perceptions of Corruption

