Team Assignment 2

Ruth Mullin, Hawa Barry, Foqia Shahid

10/21/2020

## Initial setup

library(tidyverse)

## ── Attaching packages ──────────────────────────────────────────────────────────────── tidyverse 1.3.0 ──

## ✓ ggplot2 3.2.1 ✓ purrr 0.3.3  
## ✓ tibble 3.0.0 ✓ dplyr 1.0.2  
## ✓ tidyr 1.0.2 ✓ stringr 1.4.0  
## ✓ readr 1.3.1 ✓ forcats 0.4.0

## ── Conflicts ─────────────────────────────────────────────────────────────────── tidyverse\_conflicts() ──  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

library(dplyr)  
library(readr)  
  
envsurvey <- read\_csv("envsurvey.csv")

## Parsed with column specification:  
## cols(  
## .default = col\_double(),  
## Envlist = col\_character(),  
## Edu = col\_character(),  
## Sex = col\_character(),  
## RETH = col\_character(),  
## RaceMinor = col\_character(),  
## BAdegree = col\_character(),  
## Employed = col\_character()  
## )

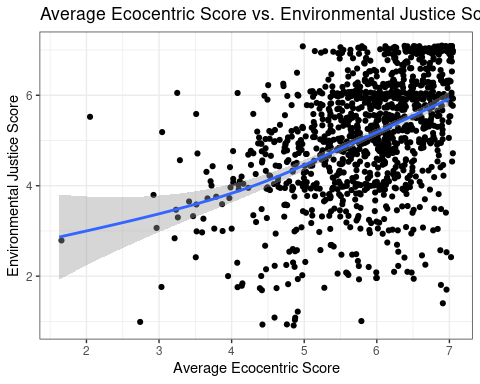
## See spec(...) for full column specifications.

## Warning: 3 parsing failures.  
## row col expected actual file  
## 1015 EnvIss\_12 a double #NULL! 'envsurvey.csv'  
## 1156 EnvIss\_12 a double #NULL! 'envsurvey.csv'  
## 1169 EnvIss\_3 a double #NULL! 'envsurvey.csv'

## Question 1

# Create columns for average ecocentric and average anthropocentric scores  
envsurvey <- envsurvey %>%  
 mutate(avg\_eco = (EnvIss\_1+EnvIss\_2+EnvIss\_6+EnvIss\_7+EnvIss\_8+EnvIss\_13+EnvIss\_15+EnvIss\_16)/8,  
 avg\_anthro = (EnvIss\_3+EnvIss\_4+EnvIss\_5+EnvIss\_9+EnvIss\_10+EnvIss\_11+EnvIss\_12+EnvIss\_14+  
 EnvIss\_17+EnvIss\_18)/10) %>%  
 drop\_na(avg\_anthro)   
  
# Plot avg\_eco against EnvJst and calculate correlation coefficient  
ggplot(data = envsurvey, aes(x = avg\_eco, y = EnvJst)) +  
 geom\_jitter() +  
 geom\_smooth() +  
 theme\_bw() +  
 labs(title = "Average Ecocentric Score vs. Environmental Justice Score",  
 x = "Average Ecocentric Score",  
 y = "Environmental Justice Score")

## `geom\_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'

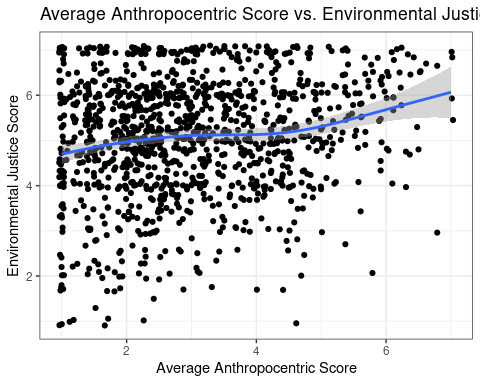


cor(envsurvey$avg\_eco, envsurvey$EnvJst)

## [1] 0.4376165

# Plot avg\_anthro against EnvJst and calculate correlation coefficient  
ggplot(data = envsurvey, aes(x = avg\_anthro, y = EnvJst)) +  
 geom\_jitter() +  
 geom\_smooth() +  
 theme\_bw() +  
 labs(title = "Average Anthropocentric Score vs. Environmental Justice Score",  
 x = "Average Anthropocentric Score",  
 y = "Environmental Justice Score")

## `geom\_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'



cor(envsurvey$avg\_anthro, envsurvey$EnvJst)

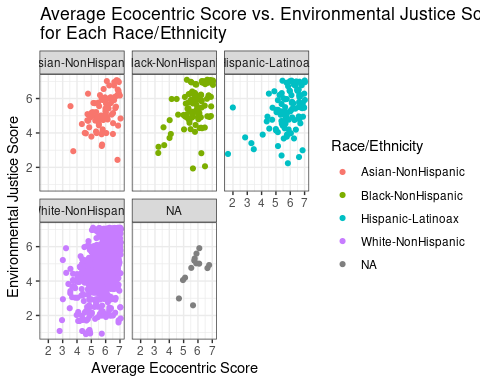
## [1] 0.1598505

evg\_eco vs EnvJst = 0.44 avg\_anthro vs EnvJst = 0.16

There is a higher correlation between average ecocentric score and environmental justice score than average anthropocentric score and environmental justice score. However, neither is very significant. This makes sense because simply believing that certain things are environmental issues does not mean that you will also believe that minorities are affected more than other groups by those issues (like the EnvJst questions ask). It makes sense that the correlation is higher for average ecocentric scores since ecocentric issues are more directly related to environmental justice. Key words such as industrial pollution are used in both the ecocentric environmental issue questions and the environmental justice questions, so people’s scores for these two types of questions could be similar.

## Question 2

# Add RETH to avg\_eco plot from question 1  
ggplot(data = envsurvey, aes(x = avg\_eco, y = EnvJst, color = RETH)) +  
 geom\_jitter() +  
 facet\_wrap(~ RETH) +  
 theme\_bw() +  
 labs(title = "Average Ecocentric Score vs. Environmental Justice Score\nfor Each Race/Ethnicity",  
 x = "Average Ecocentric Score",  
 y = "Environmental Justice Score",  
 color = "Race/Ethnicity")

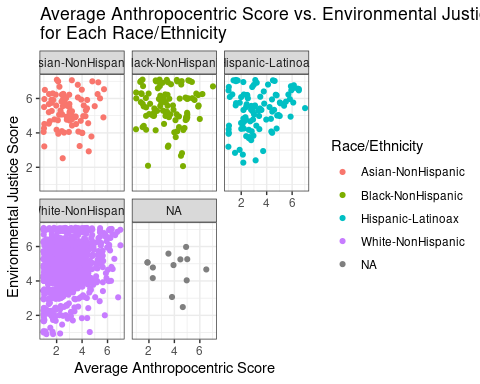


# Calculate correlation coefficient between avg\_eco and EnvJst by RETH  
envsurvey %>%  
 dplyr::group\_by(RETH) %>%  
 dplyr::summarize(correlation = cor(avg\_eco, EnvJst))

## `summarise()` ungrouping output (override with `.groups` argument)

## # A tibble: 5 x 2  
## RETH correlation  
## <chr> <dbl>  
## 1 Asian-NonHispanic 0.437  
## 2 Black-NonHispanic 0.455  
## 3 Hispanic-Latinoax 0.399  
## 4 White-NonHispanic 0.454  
## 5 <NA> 0.571

# Add RETH to avg\_anthro plot from question 1  
ggplot(data = envsurvey, aes(x = avg\_anthro, y = EnvJst, color = RETH)) +  
 geom\_jitter() +  
 facet\_wrap(~ RETH) +  
 theme\_bw() +  
 labs(title = "Average Anthropocentric Score vs. Environmental Justice Score\nfor Each Race/Ethnicity",  
 x = "Average Anthropocentric Score",  
 y = "Environmental Justice Score",  
 color = "Race/Ethnicity")



# Calculate correlation coefficient between avg\_eco and EnvJst by RETH  
envsurvey %>%  
 dplyr::group\_by(RETH) %>%  
 dplyr::summarize(correlation = cor(avg\_anthro, EnvJst))

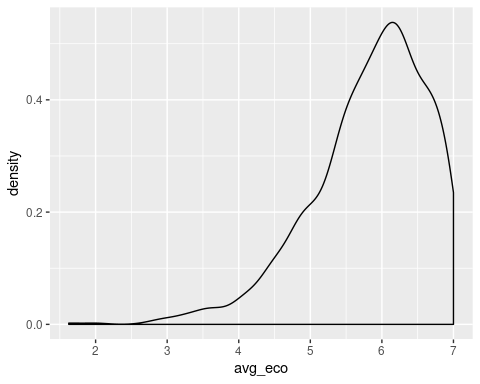
## `summarise()` ungrouping output (override with `.groups` argument)

## # A tibble: 5 x 2  
## RETH correlation  
## <chr> <dbl>  
## 1 Asian-NonHispanic 0.0519  
## 2 Black-NonHispanic -0.0413  
## 3 Hispanic-Latinoax 0.263   
## 4 White-NonHispanic 0.164   
## 5 <NA> -0.0540

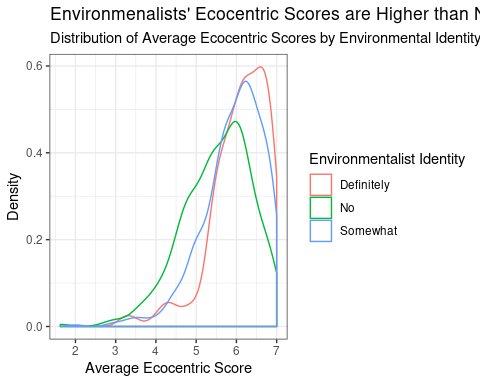
Adding in the RETH variable to our graphs for average eco/anthropocentric score vs. environmental justice score does not change our previous findings. Based on correlation coefficients for each RETH category, the findings still show the results still do not have much significance. The graphs look similar to our graphs from question 1. The ecocentric graph shows points clustering in the upper right-hand corner and the anthropocentric graph is more spread out, showing a lesser correlation.

## Question 3

# Original graph  
ggplot(envsurvey, aes(x= avg\_eco)) +  
 geom\_density()



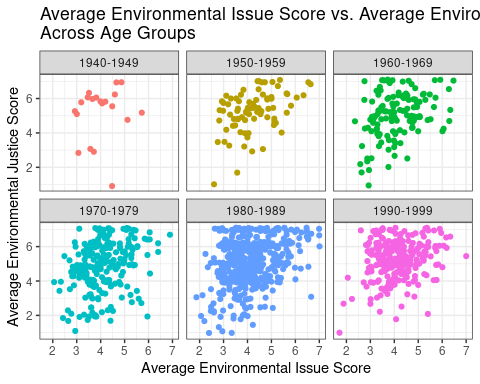
# New and improved graph  
ggplot(envsurvey, aes(x = avg\_eco, color = Envlist)) +  
 geom\_density() +  
 labs(title = "Environmenalists' Ecocentric Scores are Higher than Non-environmentalists'",   
 subtitle = "Distribution of Average Ecocentric Scores by Environmental Identity",   
 x = "Average Ecocentric Score",   
 y = "Density",   
 color = "Environmentalist Identity") +  
 theme\_bw()



## Question 4

What is the relationship between average environmental issue score and average environmental justice score at different ages?

# Group BYr by decade  
envsurvey <- envsurvey %>%  
 mutate(decade = case\_when(BYr <= 1949 ~ "1940-1949",  
 BYr <= 1959 & BYr >= 1950 ~ "1950-1959",  
 BYr <= 1969 & BYr >= 1960 ~ "1960-1969",  
 BYr <= 1979 & BYr >= 1970 ~ "1970-1979",  
 BYr <= 1989 & BYr >= 1980 ~ "1980-1989",  
 BYr <= 1999 & BYr >= 1990 ~ "1990-1999"))  
  
#  
ggplot(envsurvey, aes(x = EnvIss, y = EnvJst, color = decade)) +  
 geom\_jitter() +  
 facet\_wrap(~ decade) +  
 theme\_bw() +  
 theme(legend.position = "none") +  
 labs(title = "Average Environmental Issue Score vs. Average Environmental Justice Score\nAcross Age Groups",  
 x = "Average Environmental Issue Score",  
 y = "Average Environmental Justice Score")



#  
envsurvey %>%  
 dplyr::group\_by(decade) %>%  
 dplyr::summarize(correlation = cor(EnvJst, EnvIss))

## `summarise()` ungrouping output (override with `.groups` argument)

## # A tibble: 6 x 2  
## decade correlation  
## <chr> <dbl>  
## 1 1940-1949 0.156  
## 2 1950-1959 0.430  
## 3 1960-1969 0.346  
## 4 1970-1979 0.289  
## 5 1980-1989 0.323  
## 6 1990-1999 0.228

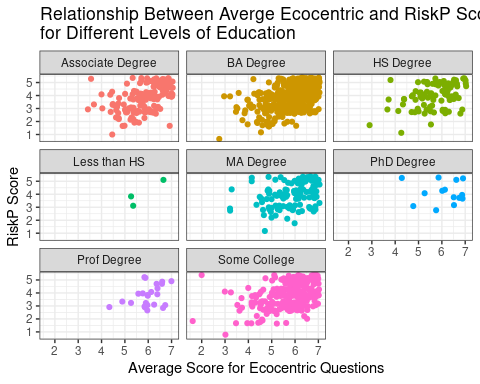
There is a positive weak association between environmental justice issues and environmental issues across all age groups mentioned in the dataset. People who care more about environmental justice issues also care more strongly about environmental issues and vice versa. People born in the fifties (1950-1959) experience the strongest association based on all other age groups with a correlation of ~0.4297688, however, the association between Environmental Justice and Environmental Issues is moderately positive.

It is important to note that the younger age groups (people from the latest 3 decades) account for more than half of the people who were surveyed. Smaller samples from the earlier 3 decades may not represent their age groups very accurately. For example, only 20 people born between 1940 and 1949 were surveyed.

## Question 5

What is the relationship between average ecocentric score and RiskP score at different levels of education?

ggplot(envsurvey, aes(x = avg\_eco, y = RiskP, color = Edu)) +  
 geom\_jitter() +  
 theme\_bw() +  
 labs(title = "Relationship Between Averge Ecocentric and RiskP Scores\nfor Different Levels of Education",  
 x = "Average Score for Ecocentric Questions",  
 y = "RiskP Score") +  
 facet\_wrap(~Edu) +  
 theme(legend.position = "none")



envsurvey %>%  
 dplyr::group\_by(Edu) %>%  
 dplyr::summarize(correlation = cor(avg\_eco, RiskP))

## `summarise()` ungrouping output (override with `.groups` argument)

## # A tibble: 8 x 2  
## Edu correlation  
## <chr> <dbl>  
## 1 Associate Degree 0.515   
## 2 BA Degree 0.454   
## 3 HS Degree 0.439   
## 4 Less than HS 0.822   
## 5 MA Degree 0.337   
## 6 PhD Degree 0.0703  
## 7 Prof Degree 0.417   
## 8 Some College 0.420

From the above analysis, we can see that there is a positive relationship between Average Score for Ecocentric questions and the RiskP score. This implies that the participants who recognized environmental issues like flooding and climate change also agreed that environmental problems pose a risk to the American public. The relationship isnt strongly correlated but there is evidence to suggest that there is some correlation.

It is significant to note that most of the participants in the study were educated people and data about people with less/no education background is scarce.

It is also significant to note that participants with PhD and Prof Degrees had overall higher RiskP and Average Econcentric scores while data for individuals with HS Degree. The “some college” category had more scattered data.

## Conclusion:

With the envsurvey data set, we explored the relationship between ecocentric and anthropocentric average score and average environmental justice score. Then we added race/ethnicity to our analysis of these questions to see if anything changed. Upon doing this, we observed that nothing changed and there was still a low correlation for all race/ethnicity categories.

Next, we looked at the correlation between average environmental issue score and average environmental justice score at different age groups using the BYr (birth year) variable. We grouped the survey participants’ birth years by decade. We discovered that there is a weak positive correlation between average environmental issue score and average environmental justice score. Out of all the age groups, those born between 1950 and 1959 had the strongest correlation between the two average scores. This might be because there were only 80 participants born in this range of years.

Finally, we asked the question, “What is the relationship between average ecocentric score and RiskP score at different levels of education?” There is a weak positive correlation between average ecocentric score and RiskP score for all education levels. The level of education does not affect this corellation. Some levels of education were very underrepresented in the group of participants, such as those with less than a high school degree.

Over all, we found that looking at different demographic variables did not affect the corellation coefficients between average scores. Even though the survey claims to have diverse participants, the groups are not equally represented. Most participants were from a certain background (white, educated, or younger (born between 1970-1999)). Among all the participants, there was a wide range of scores and answers to each question.