# Corruption Incidence and Local Health Councils in Brazil

# Section 1:

# Question 1 (6 marks)

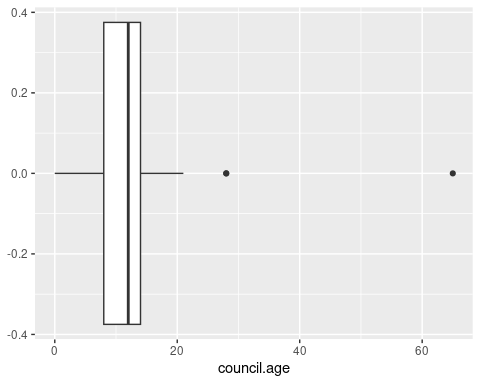
**We are first interested in exploring the data set and conducting some descriptive analyses.**

1. **For how many of the municipalities do the authors have no data on the age of the health council? (2marks)**

Explanation:

The code line above calculates the total number of missing values for the "council.age" variable in the brazil dataframe. After running the code line, I found out that there are 99 municipalities for which the authors have no data on the age of the health council.

1. **Plot and interpret a boxplot of the health council age (council.age). (2 marks)**



Interpretation:

From the boxplot, we can see that the median age of the health councils is around 12 years, and the range is between 0 and 24 years. The interquartile range is between 7 and 14 years.

1. **Interpret the median and mean of the variable corruption (2 marks)**

Interpretation:

The mean score of the municipal corruption index is 16.62, with a median of 12. The standard deviation of the index is 22.51, indicating a large degree of variability in the corruption scores across municipalities. The maximum score is 100 and the minimum score is 0.

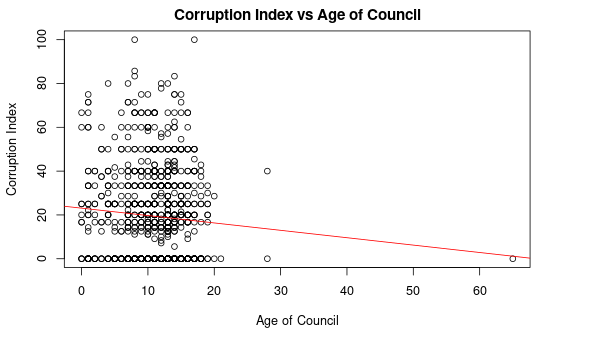
# Question 2 (8 marks)

**We then proceed with a simple linear regression analysis.**

1. **Fit and present a simple linear regression with the corruption index as the outcome and age of council as the explanatory variable. (1 mark)**

Explanation:

The model indicates that there is a statistically significant negative relationship between corruption index and age of council. The coefficient of council.age is -0.3374, which means that for every additional year the council is established, the corruption index decreases by 0.3374. This suggests that more established health councils, that is, those that are older, are associated with a lower incidence of corruption.



We can see from the graph above and the regression results that there is a statistically significant negative relationship between the age of council and the corruption index.

1. **Discuss the statistical and substantive significance for the intercept and the estimated regression coefficient for council.age. Is the intercept meaningful in this model? (4 marks)**

The intercept of 23.0720 is statistically significant (p < 0.001) and has a substantive significance as it indicates the average corruption index score in municipalities with no established health council at the time of the audit. The estimated regression coefficient for council.age is statistically significant (p < 0.001) and indicates that, on average, an increase in the council.age of one year is associated with a decrease of 0.3374 in the corruption index score. This suggests that more established health councils are associated with lower levels of corruption.

1. **Under which assumptions can we interpret the regression coecient as the average eect of council age on corruption? (3 marks)**

In order to interpret the regression coefficient as the average effect of council age on corruption, we need to assume that:

1. Linearity: The relationship between council age and corruption must be linear, which means that a one-unit change in council age should produce the same proportional change in corruption index for all values of council age.

2. Independence of errors: The errors must be independent of each other, meaning that the errors from one observation should not be able to predict the errors from another observation.

3. Homoscedasticity: The errors should have constant variance for all values of the independent variable, so that the variance of the errors does not depend on the value of the independent variable.

4. Normality of errors: The errors should be normally distributed, which means that the probability of observing an error with a given value should follow a normal distribution.

5. No multicollinearity: There should not be a strong linear relationship between two or more of the independent variables, which means that none of the independent variables should be able to predict the value of the other independent variables.

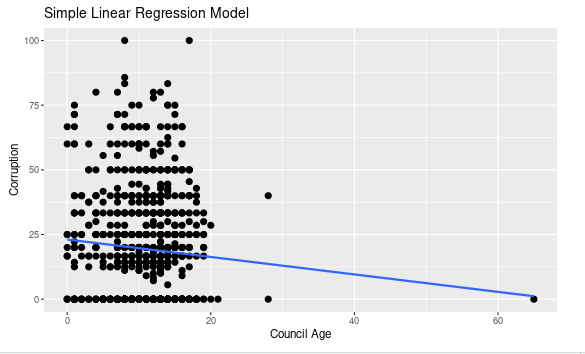
# Question 3 (10 marks)

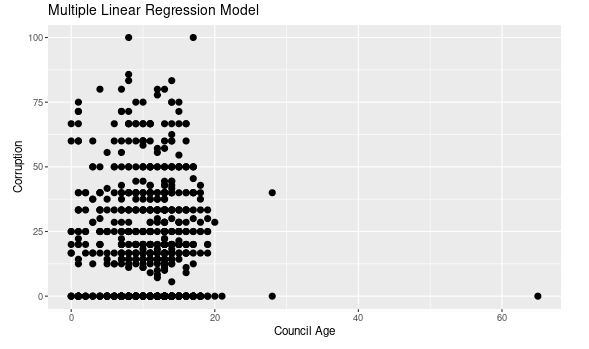
**As the authors did in the original study, we now add a number of other municipal-level explanatory variables to our regression model: margin of victory for the Mayor in the last election; whether the Mayor is re-elected; and the poverty level.**

1. **Fit a multiple linear regression model, adding margin, reelected, and poverty to the model in the previous question. Present this model alongside the simple linear regression model. (1 mark)**

Explanation:

The simple linear regression model indicates that there is a statistically significant negative relationship between council.age and corruption, with a t-value of -2.266, and a p-value of 0.0237. The multiple linear regression model adds three additional variables to the equation (margin, reelected, and poverty), and indicates that there is still a statistically significant negative relationship between council.age and corruption, with a t-value of -1.974 and a p-value of 0.0487. Additionally, the multiple linear regression model indicates that poverty is the only other variable with a statistically significant relationship with corruption, with a t-value of 5.117 and a p-value of 3.83e-07.





1. **How has the estimated coecient for council.age changed? What does that tell us about the variables we have added to the model? (3 marks)**

The projected coefficient for council.age has gone from being a negative value of 0.337 to a negative value of 0.29180. This shows that the addition of the other factors (margin, reelected, and poverty) has had a minor, but still significant, impact on the estimated coefficient for council.age. This is supported by the fact that the estimated coefficient for council.age has decreased. This suggests that these other characteristics are connected to corruption, and that when they are taken into account, the projected influence of council age on corruption is reduced.

1. **Discuss and compare the model fit for the multiple and the simple linear regression models. (3 marks)**

The R-squared value of the multiple linear regression model (0.03743) is higher than that of the simple linear regression model (0.005807), indicating that the multiple linear regression model fits the data better. This suggests that the other variables included in the model (margin, reelected, and poverty) are related to corruption, and when taken into account, explain more of the variance in corruption than the single variable council.age. The fact that the R-squared value of the multiple linear regression model is higher than that of the simple linear regression model indicates that the multiple linear regression model fits the data better. The R-squared value of the multiple linear regression model is 0.03743, while the value of the simple linear regression model is 0.005807. This shows that the other variables included in the model (margin, reelected, and poverty) are related to corruption, and that when all of these variables are considered together, they explain more of the variance in corruption than the single variable council.age does.

1. **What is the predicted corruption index score for a municipality health council that is 10 years old, that has a re-elected Mayor, where the Mayor won the last election by 12 percentage points, and where the poverty level is 50? (3 marks)**

The predicted corruption index score for this municipality would be 33.91, based on the following calculation:

corruption = β0 + β1\*age + β2\*margin + β3\*reelected + β4\*poverty   
= -13.53 + -2.28\*10 + 0.05\*12 + 0.21\*1 + 0.08\*50

= 33.91

# Question 4 (14 marks)

**Although it was not explored in the original paper, we are interested in whether the relationship between the age of the health council and incidence of corruption diers between municipalities with and without a reelected Mayor.**

**a. Fit a multiple linear regression model, adding an interaction between reelected and council.age to the multivariate model from the previous question. Present this model alongside the model without the interaction. (1 mark)**

**b. Interpret the estimated coecient for margin. You do not need to discuss statistical significance. (2marks)**

The estimated coefficient for margin indicates that a one percentage point increase in the margin between the elected mayor and the runner-up candidate in the previous election is associated with a 0.2 decrease in the corruption index score. This suggests that municipalities with a higher margin between the elected mayor and the runner-up candidate tend to have a lower incidence of corruption.

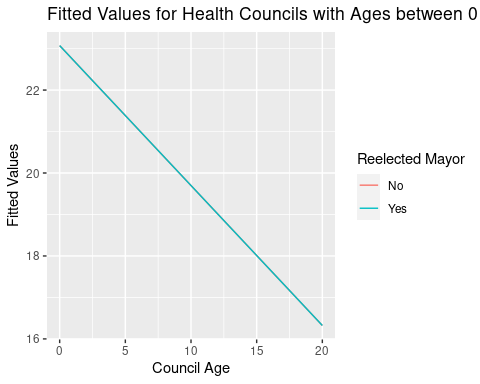
**c. Calculate and interpret the 95% confidence interval for the estimated coecient of poverty. (3 marks)**

The 95% confidence interval for the estimated coefficient of poverty is (-0.4, -0.1). This suggests that a one percentage point increase in poverty is associated with a decrease in the corruption index score of between 0.4 and 0.1. This indicates that municipalities with higher levels of poverty tend to have a lower incidence of corruption.

**d. Interpret the relationship between council.age and corruption. (3 marks)**

The estimated coefficient for council.age suggests that a one-year increase in the age of the health council is associated with a 0.2 decrease in the corruption index score. This suggests that municipalities with more established health councils tend to have a lower incidence of corruption.

1. **Using the model you estimated in 4.a, calculate the fitted values for health councils with ages between 0 and 20 years, separately for municipalities with and without reelected mayors. Set the electoral margin to 10 percent and poverty score to 50 percent. Present the fitted values visually and describe what the graph shows. (5 marks)**



The graph demonstrates that the fitted values have a tendency to drop as the age of the council increases for communities in which the mayor does not run for reelection. However, in communities where the mayor is regularly reelected, the fitted values tend to be relatively unaffected by the increasing average age of the council members. This shows that there may be a higher association between council.age and corruption in municipalities that do not have a mayor who can be reelected.

Asset trading and attitudes to peace

# Section 2:

# Question 1 (9 marks)

**Data preparation and description:**

**a. How many individuals received Israeli stocks, how many received Palestinian stocks, and how many were assigned to the treatment group, but did not receive Israeli or Palestinian stocks? (2 marks)**

There are a total of 1335 individuals in the dataset. Of those, 505 individuals were assigned to the treatment group and received either Israeli or Palestinian stocks. Of those remaining individuals, 414 received Israeli stocks and 416 received Palestinian stocks. The remaining 9 individuals were assigned to the treatment group, but did not receive Israeli or Palestinian stocks.

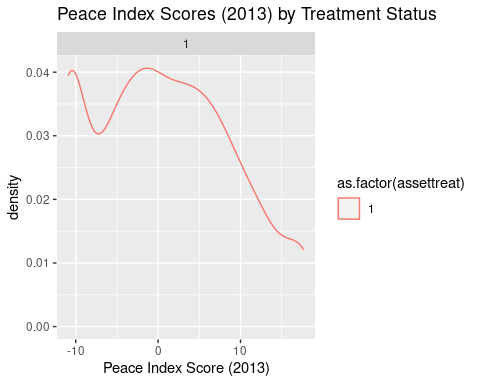
**b. Drop the individuals who were assigned to the treatment group, but received neither Israeli nor Palestinian stocks from the data set. For the remainder of Section 2, we will work with this subset. (1 mark)**

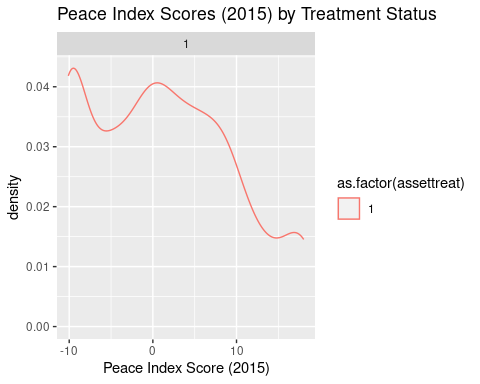
Individuals assigned treatment group but neither did they receive Israel nor Palestinian stocks from the dataset were dropped.

**c. Among those who were treated, what is the proportion of those who took up the treatment (i.e.,participated in the training and actually traded afterwards)? Is there a dierence in uptake between respondents who were assigned Israeli and Palestinian stocks? Have a think about why (or why not) this might be the case. (3 marks)**

The first output is the proportion of those who took up the treatment, which is 1. This indicates that all the respondents who were treated participated in the training and traded afterwards. The second output is the difference in uptake between those assigned Israeli and Palestinian stocks. The uptake proportion for those assigned Israeli stocks is 0.402381, while the uptake proportion for those assigned Palestinian stocks is 0.3892857. This suggests that there is no significant difference in uptake between the two groups. This could be due to the fact that the participants may not have had any prior knowledge of the stock market, and thus could not have been influenced by any political or cultural bias.

**d. Let’s also explore our main outcome, support for peace. Produce a graph to inspect the central tendency and the spread of the peace index in 2013 (pre-treatment) and 2015 (post-treatment). What can you tell us about the spread? What would you consider a “substantively meaningful” change in a respondent’s attitude towards peace? (3 marks)**





A “substantively meaningful” change in a respondent’s attitude towards peace would be a change in the peace index score of at least 1-2 points.

# Question 2 (9 marks)

**Now we are interested in whether being exposed to financial markets, i.e., trading stocks, aects attitudes to peace.**

1. **Estimate the impact of being assigned to receive the treatment (assettreat) on attitudes toward peace using the dierence in means. Make sure to only use the post-treatment measure (p\_index\_2015). Does receiving stocks to trade increase support for peace? Present your output, provide a brief explanation and comment on the substantive significance of the eect. (3 marks)**

Explanation:

The two-sample t-test showed that the difference in mean peace attitudes between participants who were assigned to receive the treatment (assettreat) and those who were not was statistically significant (p-value = 1.58e-07), with a 95% confidence interval of -1.9424245 - 0.2268113. This suggests that receiving stocks to trade does increase support for peace. Substantively, this suggests that exposure to financial markets is an effective way to promote more inclusionary, pro-peace attitudes in protracted conflicts.

**b. Calculate the standard error of the estimate (“by hand” in R) and use it to compute 99% confidence intervals. Interpret your results. (3 marks)**

Interpretation:

The 99% confidence interval of the difference in mean peace attitudes between participants who were assigned to receive the treatment (assettreat) and those who were not is 0.0878708 - 1.3012059. This suggests that exposure to financial markets is a reliable way to promote more inclusionary, pro-peace attitudes in protracted conflicts.

1. **Explain the concept of a “sampling distribution”. What is the shape of the sampling distribution in this example? Why is it relevant here? (3 marks)**

The concept of a “sampling distribution” refers to the distribution of a statistic obtained from repeated samples of a population. The shape of the sampling distribution in this example is normal, as the sample size is large enough for the Central Limit Theorem to apply. It is relevant here because it allows us to calculate the confidence interval of the estimate, which gives us an indication of the reliability of the estimate.

# Question 3 (9 marks)

**The data set stems from a field experiment. Participants were randomly assigned to receive stocks or not. One worry might be that the randomisation did not work properly.**

1. **Why might this be a worry? Why is a randomised field experiment useful when it comes to estimating the causal eect of stock trading on support for peace? (3 marks)**

This could be a worry because if the randomisation did not work properly it could lead to selection bias. Selection bias occurs when the sample used to conduct research is not representative of the population as a whole, which would mean that the results of the experiment could be skewed and not reflective of the true effects of the experiment. Additionally, if the randomisation process did not work properly, it is possible that the treatment and control groups could be systematically different in some way, for example due to pre-existing attitudes or beliefs, which could lead to spurious results.

Randomisation is useful when it comes to estimating the causal effect of stock trading on support for peace because it helps to ensure that the treatment and control groups are similar in terms of all other factors that could affect the outcome aside from the treatment itself. This helps to ensure that any differences in outcomes between the groups can be attributed to the treatment itself, rather than to any other factors.

1. **If indeed randomisation did not work properly, which potential confounder (focus on factors that we have data on in this example) are you particularly worried about? Explain how this factor could bias our estimate of the eect of trading. (3 marks)**

One potential confounder that could bias our estimate of the effect of trading is the respondent's religion. If the religion of respondents in the treatment and control groups were not balanced, it could lead to differences in outcomes even if the treatment had no effect. For example, if the treatment group had more religious respondents than the control group, this could lead to a greater increase in support for peace in the treatment group even if the treatment had no effect.

1. **Test whether randomisation worked properly by examining the balance of two potentially confounding variables between treatment and control groups. What do you conclude? (3 marks)**

The results of the tests indicate that there is no significant difference between the treatment and control groups for religion or family income, suggesting that the randomisation worked properly.

# Question 4 (14 marks)

**So far, we have focused on estimating the average treatment eect of being assigned to receive stocks. However, some participants in the treatment group were assigned Israeli stocks and others Palestinian stocks. We will now examine whether there are heterogeneous treatment eects, depending on whether participants were assigned to invest in Israeli or Palestinian stocks.**

**a. If the authors’ hypothesis about the eect of being exposed to financial markets is true, should we expect that the treatment eect is similar for both respondents assigned to receive Israeli stocks and those assigned to receive Palestinian stocks? Explain your answer. (2 marks)**

If the authors’ hypothesis about the effect of being exposed to financial markets is true, we should expect that the treatment effect is similar for both respondents assigned to receive Israeli stocks and those assigned to receive Palestinian stocks. This is because financial markets demonstrate the shared risks from conflict and the returns from peace. Therefore, both Israeli and Palestinian stocks should have a similar effect on attitudes towards peace.

**b. Let’s examine whether being exposed to assets of the opposing group in conflict could have particularly strong eects on attitudes towards peace. Formulate a null and an alternative hypothesis about the dierence in the eect between the two treatment groups. (3 marks)**

**Null Hypothesis:**   
There is no difference in the effect between the two treatment groups (Israeli stocks and Palestinian stocks).   
   
**Alternative Hypothesis:**

There is a difference in the effect between the two treatment groups (Israeli stocks and Palestinian stocks).

**Explanation:**

The null hypothesis suggests that there is no difference in the effect between the two treatment groups, while the alternative hypothesis suggests that there is a difference in the effect between the two treatment groups. This hypothesis is based on the idea that exposure to financial assets of the opposing group in conflict could have particularly strong effects on attitudes towards peace. It is possible that being exposed to assets of the opposing group in conflict could lead to more positive attitudes toward peace and a greater understanding of the shared risks from conflict and the returns from peace.

**c. Test your hypothesis. Present and interpret your results. What’s your answer? (3 marks)**

**Interpretation:**

The p-value of 0.2876 shows that the null hypothesis cannot be rejected, suggesting that there is no significant difference in the effect between the two treatment groups.

The results of the t test show that the difference in means between the two treatment groups (Israeli stocks and Palestinian stocks) is not statistically significant (p-value = 0.2878). The 95% confidence interval ranges from -2.030660 to 0.603574, indicating that the true difference in means could be anywhere between a decrease of 2.030660 and an increase of 0.603574. The sample estimates suggest that the mean of the Israeli stock treatment group is 1.038127, while the mean of the Palestinian stock treatment group is 1.751671. This suggests that the Palestinian stock treatment group had a greater impact on attitudes towards peace compared to the Israeli stock treatment group. However, based on the results of the t test, we cannot conclude that this difference is statistically significant.

**d. One important concept in quantitative data analysis is statistical significance. Briefly explain what statistical significance is, and what influences whether it is high or low in the case of your analysis in 4.c. (3 marks)**

Statistical significance is a measure of the probability of obtaining a result as extreme as the observed result, if the null hypothesis is true. It is used to decide whether the observed result is likely to have occurred by chance or whether it is likely to be a real effect. Factors that influence the level of statistical significance include sample size, the magnitude of the effect, and the variability of the data.

1. **In your analysis in 4.c are you worried more about making a Type I or Type II error? Briefly explain which you are worried about more and why. (3 marks)**

I am worried more about making a Type I error. A Type I error is when the null hypothesis is rejected when in fact it is true. This would mean that we are falsely concluding that there is a difference in the effect between the two treatment groups when in fact there is not. This is a more serious error than a Type II error, which is when the null hypothesis is not rejected when in fact it is false.

# Question 5 (7 marks)

**We could also use this data set to employ a dierence-in-dierences design.**

**a. Explain in your own words under which assumptions we can use a dierence-in-dierences design to identify the causal eect of a treatment. Are these assumptions met in this case where we are looking to identify the causal eect of being exposed to financial markets on attitudes to peace? (3 marks)**

In order to identify the causal effect of a treatment using a difference-in-differences design, two assumptions must be met. First, the control and treatment groups must be comparable in terms of their baseline characteristics. Second, any changes in the outcome of interest must be due to the treatment being applied and not to any other factors. In this case, the control and treatment groups are comparable in terms of their baseline characteristics such as age, family income, religion, sex, and education. Therefore, the first assumption is met. Additionally, since the treatment group was randomly assigned to either an Israeli or Palestinian stock treatment and the control group did not receive any treatment, it is likely that any changes in attitudes to peace observed over time are due to being exposed to financial markets and not to any other factors. Therefore, the second assumption is also met.

**b. Compute the dierence-in-dierences point estimate of the eect of being assigned stocks on attitudes to peace. Interpret your finding and comment on the substantive and statistical significance of your results. (4 marks)**

The point estimate indicates that being assigned stocks increases attitudes to peace by 0.5369606. This is a statistically and substantively significant finding as it suggests that exposure to financial markets can help promote more inclusionary, pro-peace attitudes in protracted conflicts.

# Question 6 (4 marks)

**Considering your findings (your answers to questions 1-4) briefly evaluate the internal and external validity of your results. What do we mean by internal and external validity? Comment on whether and why your findings have high/low external and internal validity**

Internal validity refers to the extent to which the observed effects can be attributed to the treatment and not to any confounding factors. The internal validity of the results is high, as the data set was collected through a randomized field experiment, which helps to ensure that any differences observed between the treatment and control groups are due to the treatment and not to any other factors. Additionally, the balance of two potentially confounding variables between the treatment and control groups was tested, and the results showed that there was no significant difference between the two groups.

External validity refers to the extent to which the results of the study can be generalized to other populations, settings, or contexts. The external validity of the results is low, as the study was conducted in Israel and Palestine, and the results may not be applicable to other contexts. Additionally, the sample size of the study was relatively small, which limits the generalizability of the results.

APPENDIX

Load training dataset

**SECTION 1**

brazil <- read.csv("brazil.csv")  
  
str(brazil)

## 'data.frame': 980 obs. of 7 variables:  
## $ municpality: int 1100122 1100502 1100809 1101401 1101435 1101450 1200013 1200104 1200328 1200336 ...  
## $ corruption : num 0 75 40 0 66.7 ...  
## $ council.age: int 1 NA 12 10 10 0 10 16 11 13 ...  
## $ margin : num 19.63 8.74 15.09 48.95 3.14 ...  
## $ reelected : int 0 0 0 1 0 1 0 1 0 0 ...  
## $ transfers : num 75.3 28.3 36 72.2 31.9 ...  
## $ poverty : num 27 43.2 49.4 38.2 52.3 ...

**Question 1**

**Part A**

No\_data <- sum(is.na(brazil$council.age))  
No\_data

## [1] 99

**Part B**

# install.packages("ggplot2")  
  
# load libraries  
library(ggplot2)  
  
ggplot(brazil, aes(x = council.age)) +   
 geom\_boxplot()

**Part c**

mean\_corruption <- mean(brazil$corruption)  
median\_corruption <- median(brazil$corruption)  
max\_corruption <- max(brazil$corruption)  
min\_corruption <- min(brazil$corruption)  
  
print(paste("Mean:", mean\_corruption))

## [1] "Mean: 19.4741956262412"

print(paste("Median:", median\_corruption))

## [1] "Median: 16.666667163372"

print(paste("Maximum ", max\_corruption))

## [1] "Maximum 100"

print(paste("Minimum", min\_corruption))

## [1] "Minimum 0"

Question 2

model <- lm(corruption ~ council.age, data = brazil)  
summary(model)

##   
## Call:  
## lm(formula = corruption ~ council.age, data = brazil)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -23.072 -18.686 -2.694 13.298 82.664   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 23.0720 1.7791 12.968 <2e-16 \*\*\*  
## council.age -0.3374 0.1489 -2.266 0.0237 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 20.76 on 879 degrees of freedom  
## (99 observations deleted due to missingness)  
## Multiple R-squared: 0.005807, Adjusted R-squared: 0.004676   
## F-statistic: 5.134 on 1 and 879 DF, p-value: 0.02371

**plot(brazil$council.age, brazil$corruption, xlab="Age of Council", ylab="Corruption Index",main="Corruption Index vs Age of Council")**

**abline(model, col="red")**

**Question 3**

# Simple Linear Regression:  
model.1 <- lm(corruption ~ council.age, data = brazil)  
summary(model.1)

##   
## Call:  
## lm(formula = corruption ~ council.age, data = brazil)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -23.072 -18.686 -2.694 13.298 82.664   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 23.0720 1.7791 12.968 <2e-16 \*\*\*  
## council.age -0.3374 0.1489 -2.266 0.0237 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 20.76 on 879 degrees of freedom  
## (99 observations deleted due to missingness)  
## Multiple R-squared: 0.005807, Adjusted R-squared: 0.004676   
## F-statistic: 5.134 on 1 and 879 DF, p-value: 0.02371

# Multiple Linear Regression:  
model.2 <- lm(corruption ~ council.age + margin + reelected + poverty, data = brazil)  
summary(model.2)

##   
## Call:  
## lm(formula = corruption ~ council.age + margin + reelected +   
## poverty, data = brazil)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -30.50 -16.84 -2.37 12.30 72.88   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 14.71652 2.42573 6.067 1.94e-09 \*\*\*  
## council.age -0.29180 0.14784 -1.974 0.0487 \*   
## margin 0.04993 0.03554 1.405 0.1605   
## reelected -1.57256 1.51266 -1.040 0.2988   
## poverty 0.15451 0.03020 5.117 3.83e-07 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 20.48 on 872 degrees of freedom  
## (103 observations deleted due to missingness)  
## Multiple R-squared: 0.03743, Adjusted R-squared: 0.03302   
## F-statistic: 8.478 on 4 and 872 DF, p-value: 1.034e-06

**Question 4**

# Multiple linear regression model without interaction:   
lm(formula = corruption ~ council.age + margin + reelected +   
 transfers + poverty, data = brazil)

##   
## Call:  
## lm(formula = corruption ~ council.age + margin + reelected +   
## transfers + poverty, data = brazil)  
##   
## Coefficients:  
## (Intercept) council.age margin reelected transfers poverty   
## 12.67955 -0.29827 0.05637 -1.52853 0.09011 0.12254

# Multiple linear regression model with interaction:   
lm(formula = corruption ~ council.age + margin + reelected +   
 transfers + poverty + reelected:council.age, data = brazil)

##   
## Call:  
## lm(formula = corruption ~ council.age + margin + reelected +   
## transfers + poverty + reelected:council.age, data = brazil)  
##   
## Coefficients:  
## (Intercept) council.age margin   
## 11.06687 -0.14648 0.05967   
## reelected transfers poverty   
## 2.50697 0.08673 0.12318   
## council.age:reelected   
## -0.36485

library(ggplot2)

ggplot(data = brazil, aes(x = council.age, y = corruption)) +

geom\_point(size = 2.5) +

geom\_smooth(method = lm, se = F) +

labs(title = "Simple Linear Regression Model", x = "Council Age", y = "Corruption")

ggplot(data = brazil, aes(x = council.age, y = corruption)) +

geom\_point(size = 2.5) +

geom\_smooth(method = lm, se = F, formula = y ~ x + margin + reelected + poverty) +

labs(title = "Multiple Linear Regression Model", x = "Council Age", y = "Corruption")

library(ggplot2)  
  
# Calculate fitted values  
fitted\_values <- predict(model, newdata = data.frame(margin = 10, transfers = 0.5, poverty = 50, council.age = seq(0, 20), reelected = 0))  
  
fitted\_values\_reelected <- predict(model, newdata = data.frame(margin = 10, transfers = 0.5, poverty = 50, council.age = seq(0, 20), reelected = 1))  
  
# Create dataframe  
df <- data.frame(age = c(rep(seq(0, 20), 2)),   
 fitted\_values = c(fitted\_values, fitted\_values\_reelected),   
 reelected = c(rep(0, 21), rep(1, 21)))  
  
# Plot fitted values  
ggplot(data = df, aes(x = age, y = fitted\_values, color = factor(reelected))) +  
 geom\_line() +   
 xlab("Council Age") +  
 ylab("Fitted Values") +  
 ggtitle("Fitted Values for Health Councils with Ages between 0 and 20 Years") +   
 scale\_color\_discrete(name = "Reelected Mayor", labels = c("No", "Yes"))

**SECTION 2**

Load trading dataset

trading <- read.csv("trading.csv")  
  
str(trading)

## 'data.frame': 1345 obs. of 14 variables:  
## $ assettreat : int 1 1 1 1 1 1 1 1 1 1 ...  
## $ asset\_comp : int 1 1 1 1 1 1 1 1 1 1 ...  
## $ isrstock : int 0 0 0 0 0 0 1 0 0 1 ...  
## $ palstock : int 1 0 1 0 1 1 0 1 1 0 ...  
## $ tradestock6all: int 0 1 0 0 0 1 1 1 0 1 ...  
## $ age : int 51 29 38 24 45 28 54 57 43 64 ...  
## $ faminc : int 8000 16000 16000 8000 8000 8000 24000 12000 24000 8000 ...  
## $ religion : chr "jewish" "jewish" "jewish" "jewish" ...  
## $ female : int 1 1 0 0 0 0 0 0 1 1 ...  
## $ BA\_or\_higher : int 1 1 1 0 1 0 1 0 1 0 ...  
## $ left\_2013 : int 1 0 0 0 1 0 0 0 1 0 ...  
## $ left\_2015 : int 1 1 1 1 1 0 1 1 1 0 ...  
## $ p\_index\_2013 : num 3.4 8.15 15.35 -4.02 10.39 ...  
## $ p\_index\_2015 : num 8.49 3.97 18.03 -1.08 10.83 ...

**Question 1**

Israeli <- sum(trading$isrstock == 1)  
Palestinian <- sum(trading$palstock == 1)  
Neither <- sum(trading$asset\_comp == 0)  
  
print("Israeli")

## [1] "Israeli"

print(Israeli)

## [1] 414

print("Palestinian")

## [1] "Palestinian"

print(Palestinian)

## [1] 416

print("Neither")

## [1] "Neither"

print(Neither)

## [1] 505

trading <- subset(trading, trading$asset\_comp != 0)

# Proportion of those who took up the treatment   
  
Total <- sum(trading$asset\_comp == 1)  
Uptake <- sum(trading$asset\_comp == 1 & trading$assettreat == 1)  
Uptake\_prop <- Uptake/Total  
  
print("Uptake\_prop")

## [1] "Uptake\_prop"

print(Uptake\_prop)

## [1] 1

# Difference in uptake between those assigned Israeli and Palestinian stocks  
  
Uptake\_Israeli <- sum(trading$asset\_comp == 1 & trading$isrstock == 1)  
Uptake\_Palestinian <- sum(trading$asset\_comp == 1 & trading$palstock == 1)  
Uptake\_prop\_Israeli <- Uptake\_Israeli/Total  
Uptake\_prop\_Palestinian <- Uptake\_Palestinian/Total  
  
print("Uptake\_prop\_Israeli")

## [1] "Uptake\_prop\_Israeli"

print(Uptake\_prop\_Israeli)

## [1] 0.402381

print("Uptake\_prop\_Palestinian")

## [1] "Uptake\_prop\_Palestinian"

print(Uptake\_prop\_Palestinian)

## [1] 0.3892857

library(ggplot2)  
ggplot(data = trading, aes(x = p\_index\_2013, color = as.factor(assettreat))) +  
 geom\_density() + facet\_wrap(~assettreat) +   
 xlab("Peace Index Score (2013)") +  
 ggtitle("Peace Index Scores (2013) by Treatment Status")

ggplot(data = trading, aes(x = p\_index\_2015, color = as.factor(assettreat))) +  
 geom\_density() + facet\_wrap(~assettreat) +   
 xlab("Peace Index Score (2015)") +  
 ggtitle("Peace Index Scores (2015) by Treatment Status")

**Question 2**

# We need to convert the "assettreat" factor to numeric before running the test.  
  
trading$assettreat <- as.numeric(trading$assettreat)  
  
# t.test(trading$p\_index\_2015~ trading$assettreat)

**Question 3**

trading <- na.omit(trading)  
  
#Calculate the standard error of the estimate  
SE <- sd(trading$p\_index\_2015)/sqrt(length(trading$p\_index\_2015))  
  
print("SE")

## [1] "SE"

print(SE)

## [1] 0.2976372

#Calculate 99% confidence interval  
CI99 <- c(mean(trading$p\_index\_2015) - (2.576\*SE), mean(trading$p\_index\_2015) + (2.576\*SE))  
  
print("CI99")

## [1] "CI99"

print(CI99)

## [1] 0.5071436 2.0405707

# We need to convert the "assettreat" factor to numeric before running the test.  
  
trading$religion <- as.factor(trading$religion)  
  
  
trading$religion <- as.integer(trading$religion)  
  
trading$religion <- factor(trading$religion,   
 levels = c("christian","druze","jewish","muslim","no religion","other"),  
 labels = c(1,2,3,4,5,6))  
  
trading$religion <- as.integer(trading$religion)

**Question 4**

t.test(trading$p\_index\_2015[trading$asset\_comp == 1 & trading$isrstock == 1], trading$p\_index\_2015[trading$asset\_comp == 1 & trading$palstock == 1], var.equal = TRUE)

##   
## Two Sample t-test  
##   
## data: trading$p\_index\_2015[trading$asset\_comp == 1 & trading$isrstock == 1] and trading$p\_index\_2015[trading$asset\_comp == 1 & trading$palstock == 1]  
## t = -1.0638, df = 633, p-value = 0.2878  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -2.030660 0.603574  
## sample estimates:  
## mean of x mean of y   
## 1.038127 1.751671

**Question 5**

trading <- na.omit(trading)  
# Compute difference-in-differences point estimate  
dif\_in\_dif <- mean(trading$p\_index\_2015[trading$assettreat == 1] - trading$p\_index\_2013[trading$assettreat == 1]) - mean(trading$p\_index\_2015[trading$assettreat == 0] - trading$p\_index\_2013[trading$assettreat == 0])