Andrew Layton

POS3703

**Final Project**

**Research Question:** What effect does an individual’s opinion about the environment have upon their choice of automobile?

**Explanation:** The link between the dependent variable, the car someone has decided to drive, and the independent variables, their opinions about the environments and their actions to conserve, are linked because it wouldn’t make sense for an environmentalist to be driving a car that harms the earth that they’re trying to save. The link here should be very strong, and show that people with more conscious opinions about the environment will drive more fuel-efficient cars. It only makes logical sense that a more fuel-efficient car would be more desirable to these individuals, barring the cost of the car itself. Even then, if they see it as a decision that will help the earth, they may decide to bite the bullet to get the greener car.

**Hypotheses:** • There is a relationship between an individual’s feelings about the environment and how often they choose to recycle when available.

• There is a relationship between an individual’s feelings about the environment and what sort of gas mileage their car gets.

• There is a relationship between an individual’s feelings about the environment and how much time they spend outside per week.

• There is a relationship between an individual’s feelings about the environment and whether they participate in any conservation projects, such as a lake clean-up or if they compost.

The null hypotheses for each of these would be that there is no relationship between an individual’s feelings about the environment and any of the independent variables.

**Unit of Analysis:** This research will be done on the individual level, and questions in this experiment will be asked to individuals.

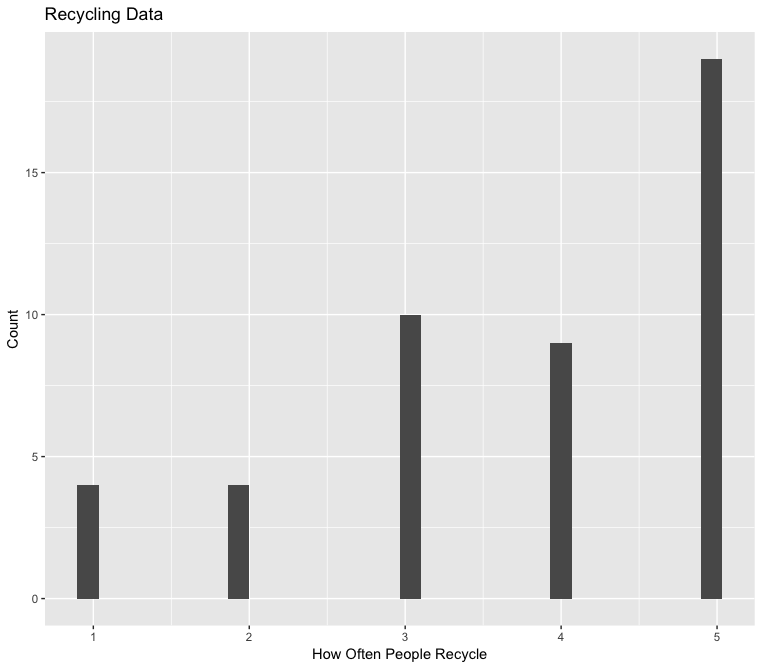
**Variables:** This research project is interested in explaining the individual’s feelings about the environment and will be used as the dependent variable. This will be done by observing the differing changes in how people go about their lives based on how high of a priority they think the environment is. The differing changes will be measured against the independent variable, which will be the mileage of the car they drive. This project will have the control variables of how often an individual recycles, how much time they spend outside per week, and if the individual participates in conservation projects. To measure the concepts of my hypotheses, I will ask participants to take a brief survey outlining each of these variables, document them in a spreadsheet, and then run analyses on them. This method has been chosen because it’ll allow me to use real data from real people that have no particular reason to not tell the truth, and will give me actual insight into how people view the environment and how they perceive their decisions in it.

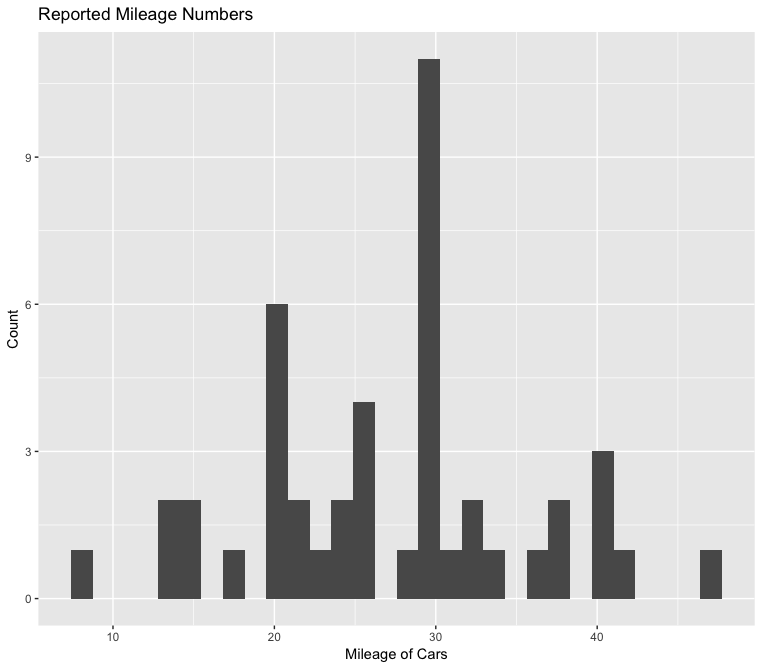
**Data:** This data was acquired via a survey that was created using Google Forms, then shared on Facebook for three days. During this time, it automatically collected and documented responses which I could then manually import into Microsoft Excel. The survey asked the following questions, in the following order: how often do you recycle; what kind of car do you drive, and if you know it, how many miles per gallon does it normally get; how many hours do you spend outside per week; do you participate in any conservation projects, such as a lake cleanup or compost; and what priority do you give to the environment in your day-to-day life. This short survey could be done in about a minute, so it solved the potential problem that people may not want to divert their time online to fill it out. Every question was mandatory, so subjects sorted themselves into different groups. Potential sources of bias in this data would be that this was only shared on Facebook, so people on different forms of social media wouldn’t really know about it. Another potential source would be that it relied on the individual to choose to do the survey, which was combated with my expression of how much it would help my project, and several friends and family members sharing it with their social circles, so that I would not be limited to just my own.

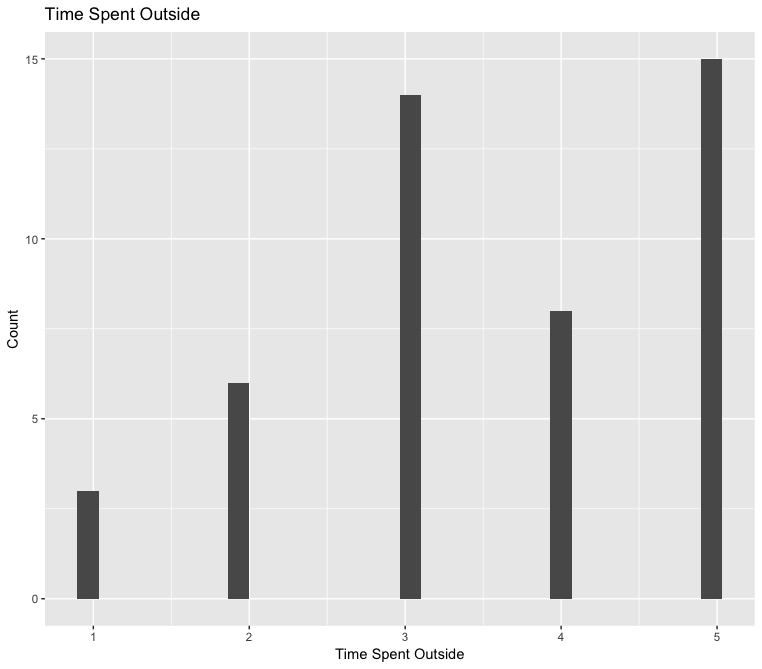
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| subject | recycle | mileage | outside | conservation | environment |
| 1 | 4 | 30 | 3 | 0 | 3 |
| 2 | 5 | 20 | 3 | 1 | 5 |
| 3 | 4 | 23 | 5 | 0 | 4 |
| 4 | 3 | 8 | 5 | 1 | 4 |
| 5 | 5 | 36 | 5 | 1 | 5 |
| 6 | 5 | 24 | 4 | 1 | 5 |
| 7 | 5 | 20 | 3 | 0 | 3 |
| 8 | 3 | 30 | 5 | 1 | 4 |
| 9 | 5 | 38 | 3 | 0 | 4 |
| 10 | 5 | 30 | 3 | 0 | 3 |
| 11 | 5 | 20 | 2 | 1 | 4 |
| 12 | 1 | 13 | 4 | 0 | 2 |
| 13 | 4 |  | 2 | 0 | 3 |
| 14 | 5 | 30 | 4 | 0 | 4 |
| 15 | 3 | 47 | 3 | 0 | 3 |
| 16 | 3 | 30 | 3 | 0 | 3 |
| 17 | 5 | 40 | 5 | 0 | 4 |
| 18 | 4 | 38 | 5 | 1 | 5 |
| 19 | 5 | 30 | 5 | 0 | 4 |
| 20 | 5 | 30 | 5 | 0 | 4 |
| 21 | 2 | 40 | 5 | 0 | 3 |
| 22 | 5 | 15 | 1 | 0 | 2 |
| 23 | 5 | 18 | 3 | 1 | 3 |
| 24 | 2 | 30 | 3 | 0 | 3 |
| 25 | 5 | 22 | 5 | 0 | 4 |
| 26 | 4 | 34 | 5 | 1 | 5 |
| 27 | 2 | 31 | 2 | 0 | 2 |
| 28 | 1 | 30 | 1 | 0 | 1 |
| 29 | 4 | 32 | 3 | 0 | 3 |
| 30 | 5 | 30 | 4 | 0 | 4 |
| 31 | 3 | 40 | 2 | 0 | 3 |
| 32 | 4 | 20 | 4 | 0 | 4 |
| 33 | 5 | 20 | 3 | 0 | 3 |
| 34 | 4 | 42 | 4 | 1 | 5 |
| 35 | 5 | 20 | 5 | 0 | 4 |
| 36 | 4 | 26 | 4 | 0 | 3 |
| 37 | 3 | 32 | 2 | 0 | 2 |
| 38 | 1 | 15 | 1 | 0 | 1 |
| 39 | 5 | 21 | 4 | 1 | 4 |
| 40 | 1 | 13 | 5 | 1 | 2 |
| 41 | 3 | 29 | 5 | 0 | 3 |
| 42 | 3 | 26 | 5 | 1 | 5 |
| 43 | 3 | 28 | 3 | 0 | 3 |
| 44 | 2 | 26 | 2 | 0 | 2 |
| 45 | 3 | 24 | 3 | 1 | 4 |
| 46 | 5 | 26 | 3 | 0 | 4 |

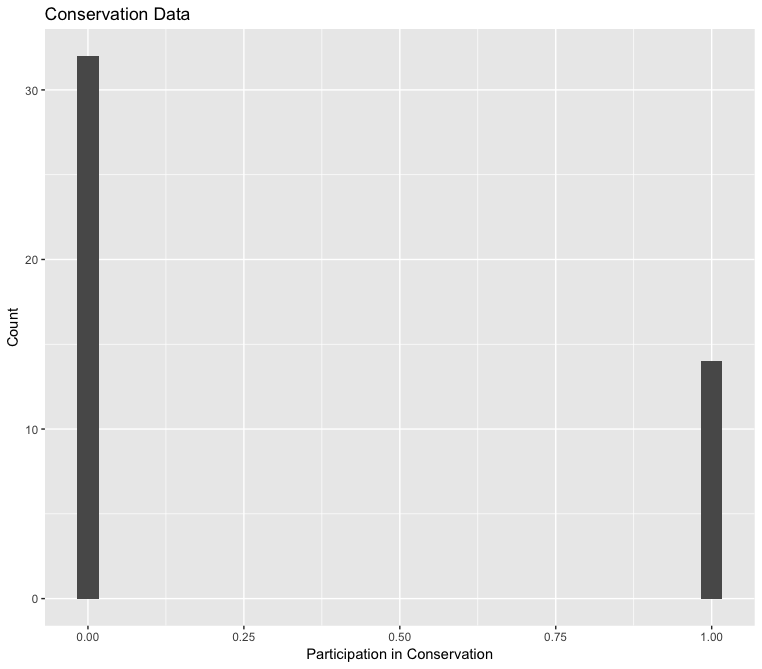
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Variable | Range | Mean | Median | Standard Deviation | Number of Observations |
| recycle | 4 (Max of 5, min of 1) | 3.761 | 4 | 1.320 | 46 |
| mileage | 39 (Max of 47, min of 8) | 27.27 | 29 | 8.497 | 45 |
| outside | 4 (Max of 5, min of 1) | 3.565 | 3.5 | 1.259 | 46 |
| conservation | 1 (Max of 1, min of 0) | 0.3043 | 0 | 0.4652 | 46 |
| environment | 4 (Max of 5, min of 1) | 3.435 | 3.5 | 1.047 | 46 |

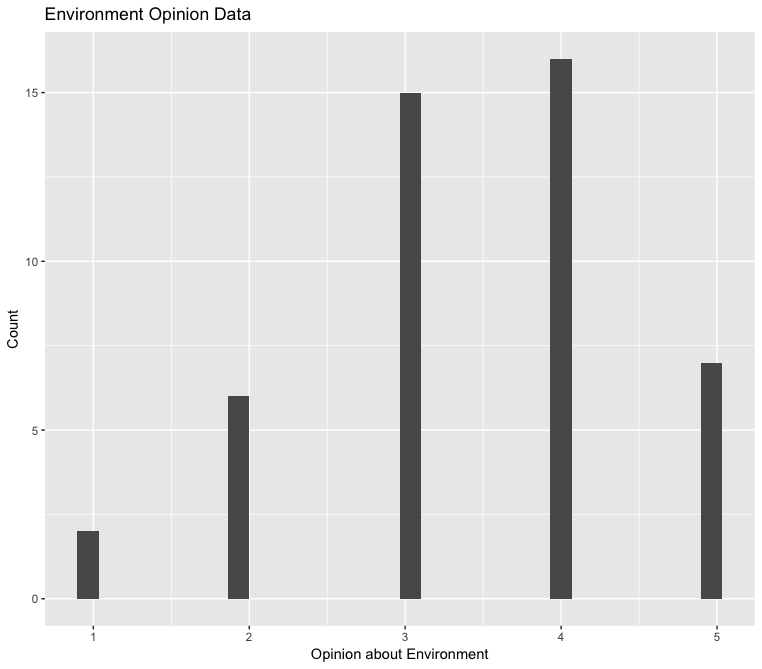
The level of measurement for each variable would be: ordinal for recycle; interval for mileage; ordinal for outside; ordinal for conservation; and ordinal for environment.

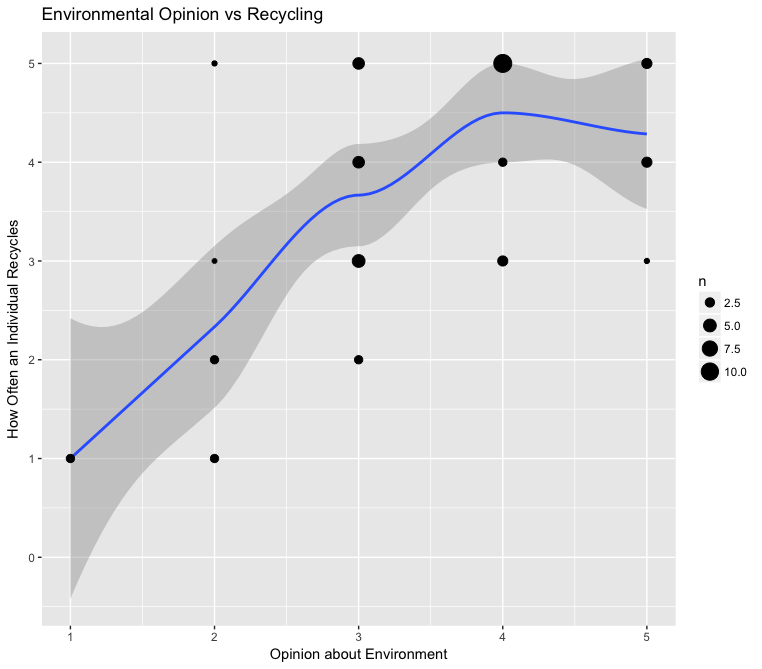






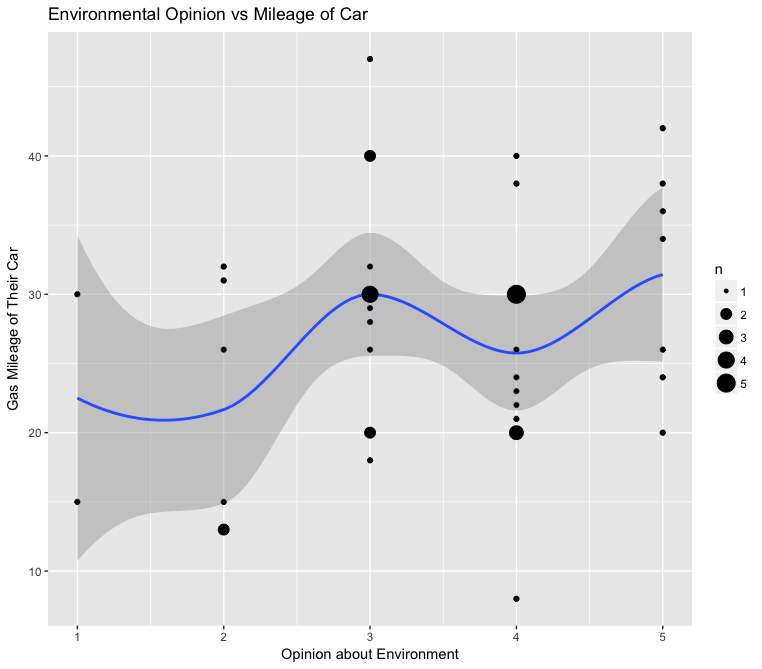






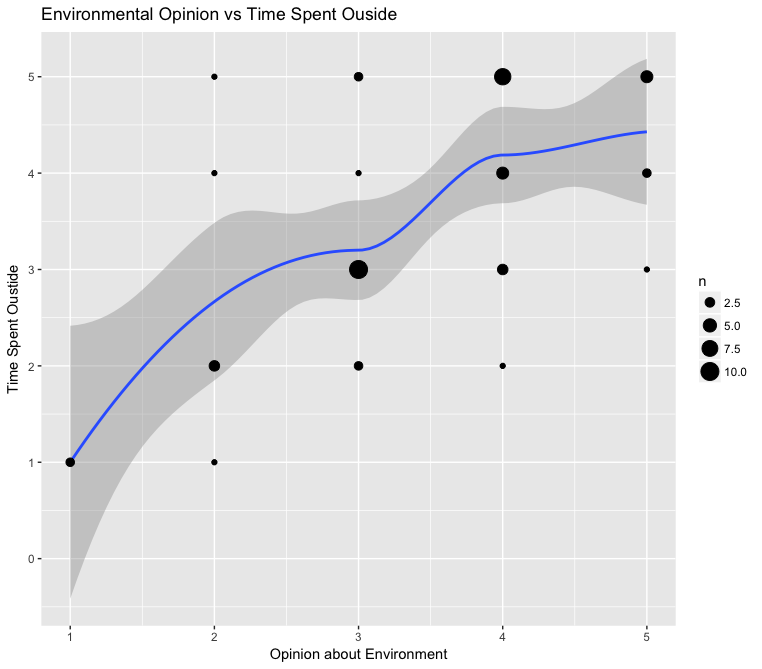
|  |  |  |  |
| --- | --- | --- | --- |
| Variable | Std. Error | t-value | P-value |
| Control | 0.5329 | 1.988 | 0.0531 |
| Recycle | 0.1485 | 5.296 | 3.61e-06 |

Pearson’s r = 0.6239, adjusted R-squared = 0.3754



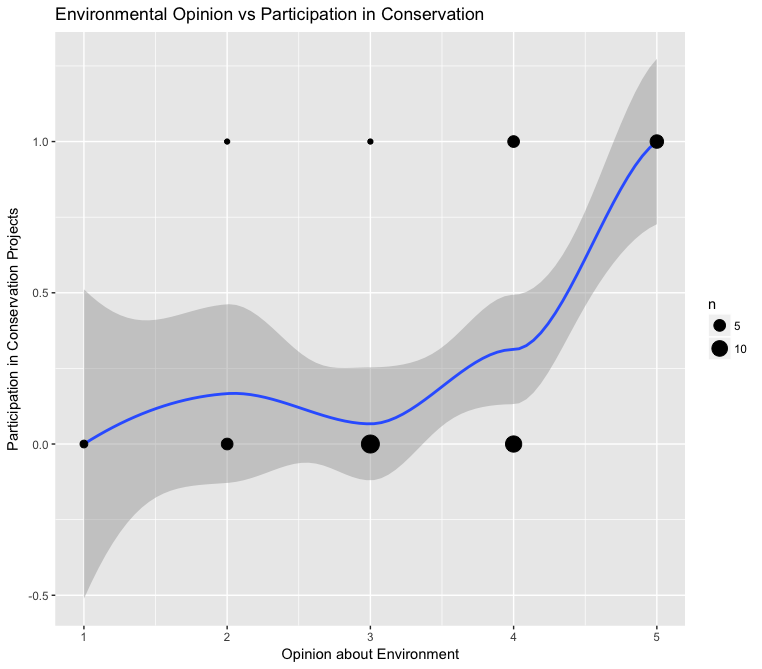
|  |  |  |  |
| --- | --- | --- | --- |
| Variable | Std. Error | t-value | P-value |
| Control | 4.307 | 4.919 | 1.32e-05 |
| Mileage | 1.197 | 1.475 | 0.148 |

Pearson’s r = 0, adjusted R-squared = 0.0260



|  |  |  |  |
| --- | --- | --- | --- |
| Variable | Std. Error | t-value | P-value |
| Control | 0.5020 | 1.871 | 0.068 |
| Outside | 0.1399 | 5.463 | 2.06e-06 |

Pearson’s r = 0.6358, adjusted R-squared = 0.3906



|  |  |  |  |
| --- | --- | --- | --- |
| Variable | Std. Error | t-value | P-value |
| Control | 0.2017 | -2.605 | 0.0125 |
| Conservation | 0.0562 | 4.297 | 9.44e-05 |

Pearson’s r = 0.5437, adjusted R-squared = 0.2795

Final regression, controlling for all variables:

|  |  |  |  |
| --- | --- | --- | --- |
| Variable | Std. Error | t-value | P-value |
| Control | 4.492 | 4.594 | 4.28e-05 |
| Environment | 7.871 | 3.369 | 0.0017 |
| Recycle | -2.829 | -2.200 | 0.0337 |
| Outside | -1.846 | 1.283 | 0.1578 |
| Conversation | 3.259 | -3.167 | 0.0030 |

Adjusted R-squared = 0.1822

**Conclusion:** Analyzing the data, we can reject the null hypothesis that there is no relationship between an individual’s opinion about the environment, how much they recycle, how much time they spend outside, and if they participate in conservation projects. Once we control all the variables in the regression, we can also extend this conclusion to the mileage of that person’s car. All of the hypotheses have been successfully confirmed.

> library(readr)

> Environment\_Study <- read\_csv("~/Documents/Environment Study.csv")

Parsed with column specification:

cols(

subject = col\_integer(),

recycle = col\_integer(),

mileage = col\_integer(),

outside = col\_integer(),

conservation = col\_integer(),

environment = col\_integer()

)

> View(Environment\_Study)

> attach(Environment\_Study)

> mean(recycle)

[1] 3.76087

> median(recycle)

[1] 4

> library("foreign", lib.loc="/Library/Frameworks/R.framework/Versions/3.3/Resources/library")

> library("gdata", lib.loc="/Library/Frameworks/R.framework/Versions/3.3/Resources/library")

gdata: read.xls support for 'XLS' (Excel 97-2004) files ENABLED.

gdata: read.xls support for 'XLSX' (Excel 2007+) files ENABLED.

Attaching package: ‘gdata’

The following object is masked from ‘package:stats’:

nobs

The following object is masked from ‘package:utils’:

object.size

The following object is masked from ‘package:base’:

startsWith

> library("ggplot2", lib.loc="/Library/Frameworks/R.framework/Versions/3.3/Resources/library")

> library("gmodels", lib.loc="/Library/Frameworks/R.framework/Versions/3.3/Resources/library")

> library("gtools", lib.loc="/Library/Frameworks/R.framework/Versions/3.3/Resources/library")

> sd(recycle)

[1] 1.319676

> max(recycle) - min(recycle)

[1] 4

> max(mileage)

[1] NA

> max(mileage, na.rm = T)

[1] 47

> min(mileage)

[1] NA

> min(mileage, na.rm = T)

[1] 8

> mean(mileage)

[1] NA

> mean(mileage, na.rm = T)

[1] 27.26667

> median(mileage, na.rm = T)

[1] 29

> sd(mileage, na.rm = T)

[1] 8.497058

> mean(outside)

[1] 3.565217

> median(outside)

[1] 3.5

> sd(outside)

[1] 1.258786

> mean(environment)

[1] 3.434783

> median(environment)

[1] 3.5

> sd(environment)

[1] 1.046734

> mean(conservation)

[1] 0.3043478

> median(conservation)

[1] 0

> sd(conservation)

[1] 0.4652151

> qplot(recycle, geom="histogram")

`stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

> ggplot(Environment\_Study, aes(recycle)) + geom\_histogram()

`stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

> qplot(recycle, geom="histogram") + labs(x = "How Often People Recycle", y = "Count", title = "Recycling Data")

`stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

> qplot(mileage, geom = "histogram")

`stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

Warning message:

Removed 1 rows containing non-finite values (stat\_bin).

> qplot(mileage, geom = "histogram") + labs(x = "Mileage of Cars", y = "Count", title = "Reported Mileage Numbers")

`stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

Warning message:

Removed 1 rows containing non-finite values (stat\_bin).

> qplot(outside, geom = "histogram") + labs("Time Spent Outside", y = "Count", title = "Time Spent Outside")

`stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

> qplot(outside, geom = "histogram") + labs(x = "Time Spent Outside", y = "Count", title = "Time Spent Outside")

`stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

> qplot(conservation, geom = "histogram") + labs(x = "Participation in Conservation", y = "Count", title = "Conservation Data")

`stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

> qplot(environment, geom = "histogram") + labs(x = "Opinion about Environment", y = "Count", title = "Environment Opinion Data")

`stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

> ggplot(Environment\_Study, aes(environment, recycle)) + geom\_point()

> ggplot(Environment\_Study, aes(environment, recycle)) + geom\_point() + geom\_smooth()

> ggplot(Environment\_Study, aes(environment, recycle)) + geom\_point() + geom\_smooth() + geom\_count()

> ggplot(Environment\_Study, aes(environment, recycle)) + geom\_point() + geom\_smooth() + geom\_count() + labs(x = "Opinion about Environment", y = "How Often an Individual Recycles", title = "Environmental Opinion vs Recycling")

> t.test(recycle,environment)

Welch Two Sample t-test

data: recycle and environment

t = 1.313, df = 85.566, p-value = 0.1927

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-0.1676543 0.8198282

sample estimates:

mean of x mean of y

3.760870 3.434783

> reg = lm(recycle ~ environment)

> summary(reg)

Call:

lm(formula = recycle ~ environment)

Residuals:

Min 1Q Median 3Q Max

-1.99206 -0.84568 0.00794 0.79453 2.36772

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 1.0591 0.5329 1.988 0.0531 .

environment 0.7866 0.1485 5.296 3.61e-06 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.043 on 44 degrees of freedom

Multiple R-squared: 0.3893, Adjusted R-squared: 0.3754

F-statistic: 28.04 on 1 and 44 DF, p-value: 3.611e-06

> cor(recycle, environment)

[1] 0.6239082

> reg = lm(mileage ~ environment)

> summary(reg)

Call:

lm(formula = mileage ~ environment)

Residuals:

Min 1Q Median 3Q Max

-20.247 -6.482 1.753 5.518 20.518

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 21.188 4.307 4.919 1.32e-05 \*\*\*

environment 1.765 1.197 1.475 0.148

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 8.386 on 43 degrees of freedom

(1 observation deleted due to missingness)

Multiple R-squared: 0.04814, Adjusted R-squared: 0.02601

F-statistic: 2.175 on 1 and 43 DF, p-value: 0.1476

> ggplot(Environment\_Study, aes(environment, mileage)) + geom\_point() + geom\_smooth() + geom\_count() + labs(x = "Opinion about Environment", y = "Gas Mileage of Their Car", title = "Environmental Opinion vs Mileage of Car")

`geom\_smooth()` using method = 'loess'

There were 11 warnings (use warnings() to see them)

> cor(mileage, environment)

[1] NA

> cor(mileage, environment, na.rm = T)

Error in cor(mileage, environment, na.rm = T) :

unused argument (na.rm = T)

> cor(mileage, environment)

[1] NA

> ggplot(Environment\_Study, aes(environment, outside)) + geom\_point() + geom\_smooth() + geom\_count() + labs(x = "Opinion about Environment", y = "Time Spent Oustide", title = "Environmental Opinion vs Time Spent Ouside")

> cor(environment, outside)

[1] 0.6357563

> reg = lm(outside ~ environment)

> summary(reg)

Call:

lm(formula = outside ~ environment)

Residuals:

Min 1Q Median 3Q Max

-1.9973 -0.6448 -0.2328 0.9438 2.5318

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 0.9392 0.5020 1.871 0.068 .

environment 0.7646 0.1399 5.463 2.06e-06 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.9826 on 44 degrees of freedom

Multiple R-squared: 0.4042, Adjusted R-squared: 0.3906

F-statistic: 29.85 on 1 and 44 DF, p-value: 2.06e-06

> ggplot(Environment\_Study, aes(environment, conservation)) + geom\_point() + geom\_smooth() + geom\_count() + labs(x = "Opinion about Environment", y = "Participation in Conservation Projects", title = "Environmental Opinion vs Participation in Conservation")

> cor(environment, conservation)

[1] 0.5436508

> reg = lm(conservation ~ environment)

> summary(reg)

Call:

lm(formula = conservation ~ environment)

Residuals:

Min 1Q Median 3Q Max

-0.4409 -0.1993 -0.1993 0.3175 1.0423

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -0.52557 0.20174 -2.605 0.0125 \*

environment 0.24162 0.05624 4.297 9.44e-05 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3949 on 44 degrees of freedom

Multiple R-squared: 0.2956, Adjusted R-squared: 0.2795

F-statistic: 18.46 on 1 and 44 DF, p-value: 9.437e-05

> reg = lm(mileage~environment + recycle + outside + conservation)

> summary(reg)

Call:

lm(formula = mileage ~ environment + recycle + outside + conservation)

Residuals:

Min 1Q Median 3Q Max

-16.0809 -4.5676 -0.2247 5.4324 16.7753

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 20.636 4.492 4.594 4.28e-05 \*\*\*

environment 7.871 2.336 3.369 0.00168 \*\*

recycle -2.829 1.286 -2.200 0.03365 \*

outside -1.846 1.283 -1.439 0.15799

conservation -10.322 3.259 -3.167 0.00295 \*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7.684 on 40 degrees of freedom

(1 observation deleted due to missingness)

Multiple R-squared: 0.2565, Adjusted R-squared: 0.1822

F-statistic: 3.45 on 4 and 40 DF, p-value: 0.01632