

Working_with_Data

Table of contents

0.1 Packages & library	1
0.2 Load	1
0.3 Variables	1
0.4 Bivariate Analysis	5

0.1 Packages & library

```
library(tidyverse)
```

```
-- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
v dplyr      1.1.4      v readr      2.1.5
v forcats    1.0.0      v stringr    1.5.1
v ggplot2    3.5.1      v tibble     3.2.1
v lubridate  1.9.4      v tidyr      1.3.1
v purrr      1.0.4
-- Conflicts ----- tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()     masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(descr)
library(knitr)
library(dplyr)
library(Hmisc)
```

Attaching package: 'Hmisc'

The following objects are masked from 'package:dplyr':

src, summarize

The following objects are masked from 'package:base':

format.pval, units

```
library(readr)
library(readxl)
library(ggplot2)
```

0.2 Load

```
Project_Data = read.csv("/cloud/project/Data/Connection_to_Nature_Data.csv",
header = TRUE)
```

0.3 Variables

```
# VARIABLE 1: People's Age

# Age 18 and above

Project_Data <- subset(Project_Data, D_Age >= 18)

# Group D_Age into 4 age range's and Labeling

Project_Data$Age_Group <- cut(Project_Data$D_Age,
                              breaks = c(18, 25, 40, 65, Inf),
                              labels = c("18-25", "26-40", "40-65", "65+"),
                              right = TRUE,
                              include.lowest = TRUE)

# Frequency table

freq(as.ordered(Project_Data$Age_Group), plot = FALSE)
```

```
as.ordered(Project_Data$Age_Group)
      Frequency Percent Cum Percent
18-25         49   9.515      9.515
26-40         85  16.505     26.019
40-65        301  58.447     84.466
65+          80  15.534    100.000
Total        515 100.000
```

I choose this variable (age) because I think it would be important to look at in reference to how loneliness and time spent in nature varies among age groups. Perhaps depending on the age group, there will be more positive benefits to those exposed to nature in relation to loneliness.

```
# VARIABLE 2: How many hours on average do you currently spend in nature per week?

# Group D_hours into 4 categories & Labeling

Project_Data$Nature_Hours_Group <- cut(Project_Data$D_hours,
                                       breaks = c(0, 5, 15, 30, Inf),
                                       labels = c("Low (0-5)", "Moderate (6-15)", "High (16-30)", "Very High (30+)"),
                                       include.lowest = TRUE)

# Frequency table

freq(as.ordered(Project_Data$Nature_Hours_Group), plot = FALSE)
```

```
as.ordered(Project_Data$Nature_Hours_Group)
      Frequency Percent Cum Percent
Low (0-5)       137  26.602     26.60
Moderate (6-15)  230  44.660     71.26
High (16-30)    118  22.913     94.17
Very High (30+)   30   5.825    100.00
Total          515 100.000
```

This is very important. This variable (hours spent in nature) is important because when I did the literature review assignment, depending on the time spent in nature, actually lowered both social loneliness and emotional loneliness, but it depends how much time was spent in nature.

```
# VARIABLE 3: People's experience a general sense of emptiness (survey response)

# Labeling

Project_Data$Lon_1 <- factor(Project_Data$Lon_1,
                              levels = c(1, 2, 3),
                              labels = c("yes", "more or less", "no"))

# Frequency table

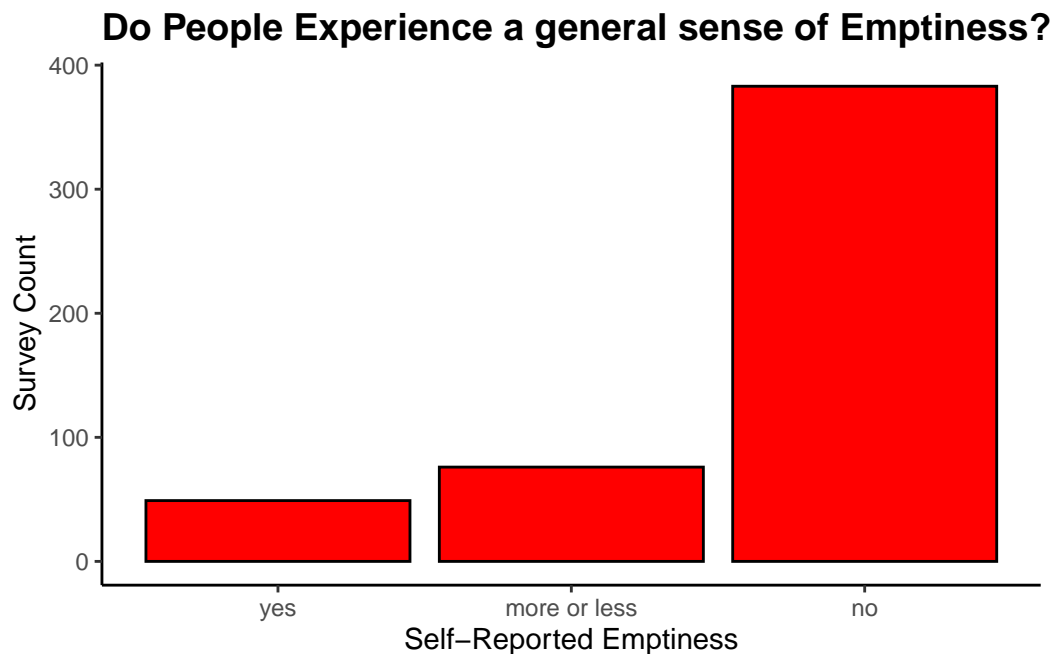
freq(as.ordered(Project_Data$Lon_1), plot = FALSE)
```

```
as.ordered(Project_Data$Lon_1)
```

	Frequency	Percent	Valid Percent	Cum Percent
yes	49	9.515	9.646	9.646
more or less	76	14.757	14.961	24.606
no	383	74.369	75.394	100.000
NA's	7	1.359		
Total	515	100.000	100.000	

```
# Bar graph

ggplot(data = subset(Project_Data, !is.na(Lon_1)), aes(x = Lon_1)) +
  geom_bar(fill = "red", color = "black") +
  xlab("Self-Reported Emptiness") +
  ylab("Survey Count") +
  ggtitle("Do People Experience a general sense of Emptiness?") +
  theme_classic() +
  theme(plot.title = element_text(size = 14, face = "bold"))
```



- Lit Review Assign: I choose this variable (people's sense of emptiness) because this can be a reason for social/emotional loneliness. If time spent nature is associated with lower loneliness on these two paths, we might also see a decrease in emptiness to those who spend more time in nature.
- Univariate Data Visualization Assign: The first graph illustrates people's experiences of a general sense of emptiness. Overall, most respondents said no, while the fewest number of respondents said yes.

```
# VARIABLE 4: I miss having people around (survey response)

# Labeling

Project_Data$Lon_4 <- factor(Project_Data$Lon_4,
                             levels = c(1, 2, 3),
                             labels = c("yes", "more or less", "no"))

# Frequency table

freq(as.ordered(Project_Data$Lon_4), plot = FALSE)
```

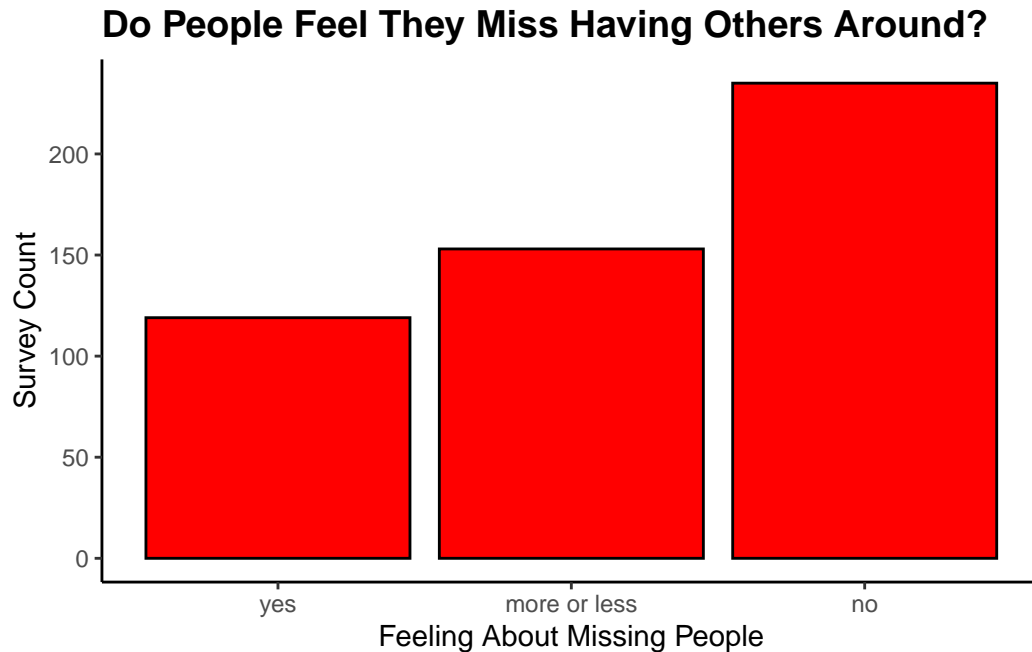
```
as.ordered(Project_Data$Lon_4)
```

	Frequency	Percent	Valid Percent	Cum Percent
yes	119	23.107	23.47	23.47
more or less	153	29.709	30.18	53.65
no	235	45.631	46.35	100.00

NA's	8	1.553	
Total	515	100.000	100.00

```
# Bar graph

ggplot(data = subset(Project_Data, !is.na(Lon_4)), aes(x = Lon_4)) +
  geom_bar(fill = "red", color = "black") +
  xlab("Feeling About Missing People") +
  ylab("Survey Count") +
  ggtitle("Do People Feel They Miss Having Others Around?") +
  theme_classic() +
  theme(plot.title = element_text(size = 14, face = "bold"))
```



- Lit Review Assign: This variable (missing social interaction) could be important because social loneliness is being examined here. Comparing this to time spent in nature can help show whether nature can also regulate/help social loneliness as well.
- Univariate Data Visualization Assign: The second graph illustrates people's survey responses to whether individuals feel that they miss others in their lives. This is similar to the first graph. Most respondents said no, while the fewest number of respondents said yes.

```
# VARIABLE 5: I have high self-esteem (survey response)

# Labeling

Project_Data$SE_1 <- factor(Project_Data$SE_1,
  levels = c(1, 2, 3, 4, 5),
  labels = c("not very true of me", "2", "3", "4", "very true of me"),
  ordered = TRUE)

# Frequency table

freq(as.ordered(Project_Data$SE_1), plot = FALSE)
```

```
as.ordered(Project_Data$SE_1)
      Frequency Percent Cum Percent
not very true of me      35   6.796    6.796
2                      69  13.398   20.194
3                     146  28.350   48.544
4                     193  37.476   86.019
very true of me         72  13.981  100.000
Total                   515 100.000
```

Lastly, I also choose this variable (people's self-esteem) because those who experience loneliness and spend little time in nature differ from those who don't feel loneliness and do spend time in nature. Perhaps those who do spend more time have higher levels of agreement to self-esteem compared to those who do not.

0.4 Bivariate Analysis

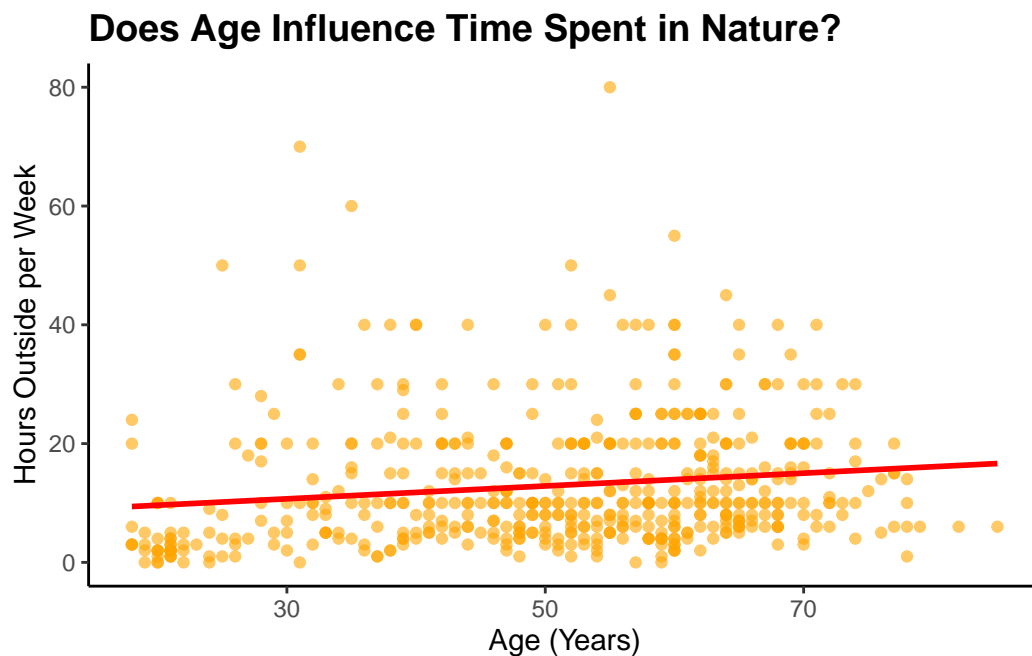
```
# Bivariate 1: Correlation between age and time spent in nature

Project_Data <- subset(Project_Data, !is.na(D_Age) & !is.na(D_hours))

# Scatter plot

ggplot(Project_Data, aes(x = D_Age, y = D_hours)) +
  geom_point(color = "orange", alpha = 0.6) +
  geom_smooth(method = "lm", se = FALSE, color = "red") +
  xlab("Age (Years)") +
  ylab("Hours Outside per Week") +
  ggtitle("Does Age Influence Time Spent in Nature?") +
  theme_classic() +
  theme(plot.title = element_text(size = 14, face = "bold"))
```

`geom_smooth()` using formula = 'y ~ x'



- The scatter plot shown here shows that there is a slight positive correlation between age and time spent outside hourly per week. The positive correlation shows that older individuals might spend a little bit more time in nature. However, when it comes to my research interest concerning the relationship between loneliness and nature, age does not seem to be a strong factor.

```
# Bivariate 2: Correlation between time spent in nature and general emptiness

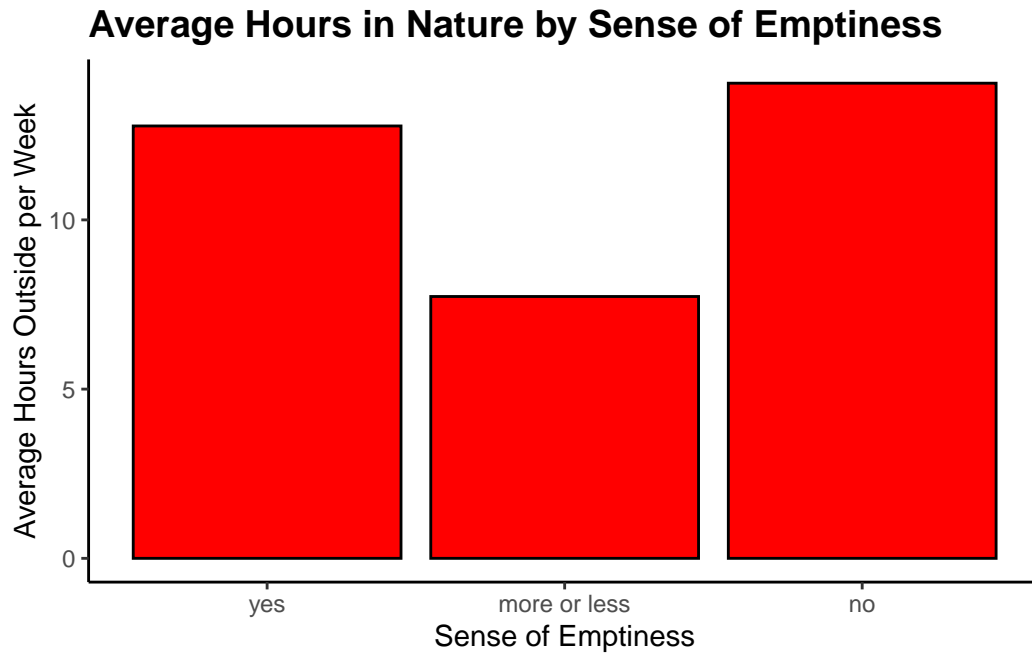
summary_data <- Project_Data %>%
  filter(!is.na(Lon_1) & !is.na(D_hours)) %>%
  group_by(Lon_1) %>%
  summarise(Avg_Hours = mean(D_hours))

# Bar graph

ggplot(summary_data, aes(x = Lon_1, y = Avg_Hours, fill = Lon_1)) +
  geom_bar(stat = "identity", fill = "red", color = "black") +
  xlab("Sense of Emptiness") +
```

```
ylab("Average Hours Outside per Week") +
ggtitle("Average Hours in Nature by Sense of Emptiness") +
theme_classic() +
theme(plot.title = element_text(size = 14, face = "bold")) +
guides(fill = FALSE)
```

Warning: The `<scale>` argument of `guides()` cannot be `FALSE`. Use "none" instead as of ggplot2 3.3.4.



- The bar graph shown here illustrates the relationship between people's sense of emptiness, and their average hours spent in nature. Individuals who say YES to feeling empty, show, a high number of hours outside, and are like those who say NO to feeling empty as well. However, people who said MORE OR LESS, spent the least amount of time outside. They would mean that the relationship between a sense of emptiness and time spent in nature is not straightforward and there could be other factors influencing nature and loneliness.