Data Design Difference Challenge 2023

Jordan Rodriguez

2023-11-11

FOOD. What Works? What Doesn't?

There are many aspects of our everyday lives that are influenced by food, as this year's Data Design Challenge prompt describes. When I read this year's prompt, I immediately gravitated towards the idea of exploring food waste & sustainability through the lens of data analysis. As an avid gardener, certified master composter, and environmental scientist by training, I felt up for this task! Now, I am a second-year biology PhD student in the machine learning/population genetics Kern-Ralph Co-Lab, interested in computational methods and tools development, and I enjoy data analysis and visualization. To this end, I wanted to create something that gave viewers a reason to view! So, I decided on a bandersnatch-style "choose-your-adventure" poster, challenging viewers on their knowledge of food waste on campus, hopefully motivating change in their personal lives as they navigate through my interactive infographic.

I also believe in open-source science so I decided to write this document as a source of supplemental material to my competition submission. In this document, you will discover how each figure was made, and see where all of the data comes from. I made this document using Rmarkdown

---

make it to the end of this document for the answer to this joke!

What did the composting rapper say?

R version & os info

```
version
```

```
##                _
## platform       aarch64-apple-darwin20
## arch           aarch64
## os             darwin20
## system         aarch64, darwin20
## status
## major          4
## minor          2.3
## year           2023
## month          03
## day            15
## svn rev        83980
## language       R
## version.string R version 4.2.3 (2023-03-15)
## nickname       Shortstop Beagle
```

Data used in this exercise

The cleaned data used for this project can be found in the same repository as this document, and the raw data can be found at this **link** and this **link**

My approach was to download the excel files available on the UO website, subset the data to create the .tsv (I used BBedit on Mac to do this) files and read them into Rstudio. I used the Lane County Waste Data document as a reference for calculating Food Waste per student and as a comparison from 2007 estimates.

Reading in Waste Data

```
# Reading in the waste data
waste <- read.csv("uowastedata.tsv", sep='\t', header = FALSE)
waste <- as.data.frame(t(waste))
names(waste) <- as.matrix(waste[1, ])
waste <- waste[-1, ]
waste[] <- lapply(waste, function(x) type.convert(as.character(x)))
waste
```

```
##     Year All_other Paper Yard_Debris Glass_Metal_Plastic Food_materials
## V2  1993     148.3 632.7        12.7                43.8           19.2
## V3  1994     151.4 625.5       125.5                62.8           23.4
## V4  1995      59.8 585.5       185.4                70.2           26.2
## V5  1996      92.5 692.3       228.1                69.7           35.6
## V6  1997     114.3 751.1       150.8                93.8           34.2
## V7  1998     141.9 810.5        86.0                71.5           35.0
## V8  1999     120.2 837.4       188.0                85.0           47.3
## V9  2000     246.9 809.7       134.5                82.3           45.8
## V10 2001     103.6 810.3       135.2               138.3           18.1
## V11 2002     120.4 762.9       179.8               145.1           21.1
## V12 2003     124.6 849.4        60.0               132.8           27.2
## V13 2004     120.4 762.9       179.8               145.1           21.1
## V14 2005     118.4 850.8       359.4               135.0           48.0
## V15 2006     135.3 873.3       235.0               128.2           50.4
## V16 2007     203.9 832.6       256.6               152.0           49.4
## V17 2008     219.1 842.3       239.0               171.4           64.5
## V18 2009     182.9 743.2       282.0               151.5           91.3
## V19 2010     209.4 732.8       295.0               130.6          121.7
## V20 2011     243.3 748.8       334.0               137.8          139.5
## V21 2012     182.7 693.8       219.5               145.4          188.2
## V22 2013     209.7 611.8       234.5               127.8          312.4
## V23 2014     214.0 571.0       466.9               118.7          364.1
## V24 2015     224.9 672.1       580.8               113.6          404.3
## V25 2016     235.7 573.9       416.8               117.2          412.2
## V26 2017     235.0 582.6       435.8               124.6          453.7
## V27 2018     277.6 507.2       372.0               126.6          436.0
## V28 2019     955.8 569.8       387.0               107.9          388.7
## V29 2020     336.9 408.7       372.0                42.3          227.3
## V30 2021     247.5 243.9       372.0                26.8          114.2
##     Landfill
## V2    1914.5
## V3    1759.8
## V4    1641.5
## V5    1562.8
## V6    1555.9
```

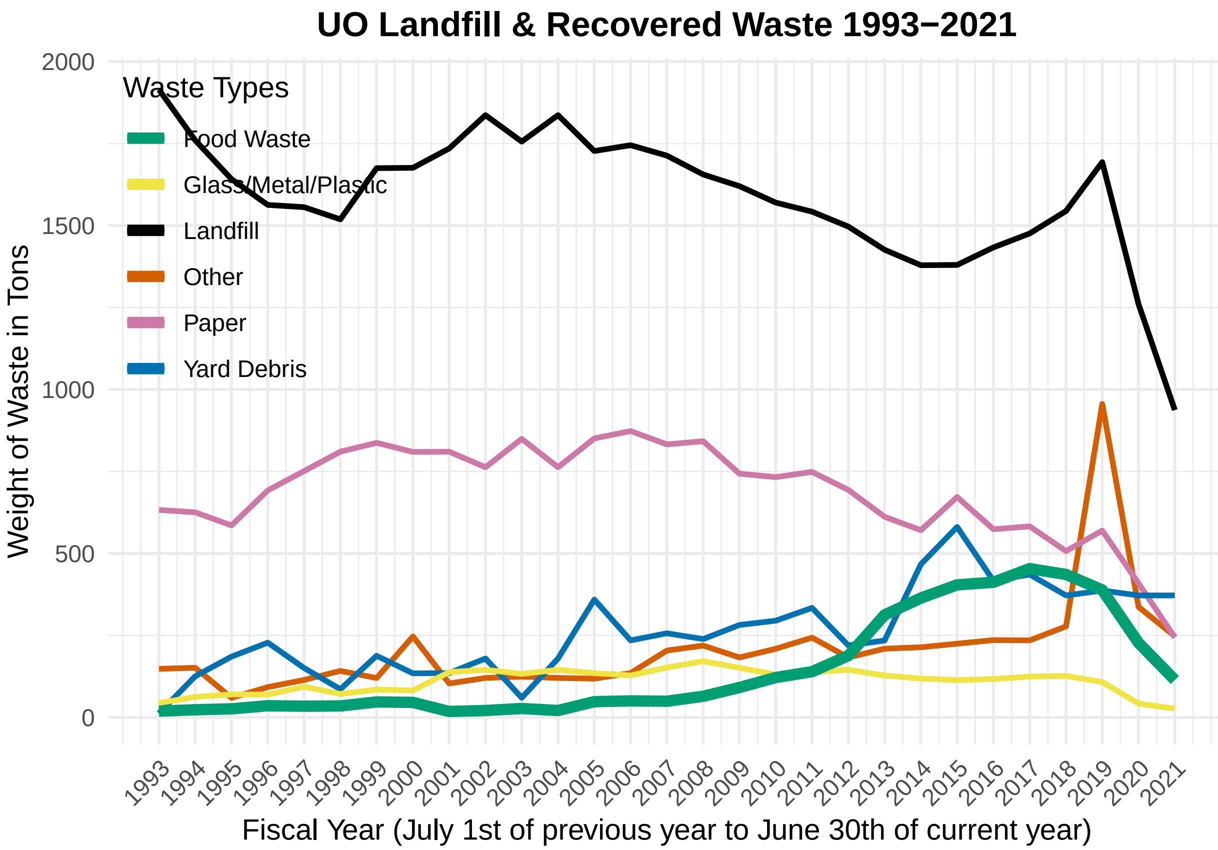
```
## V7    1518.8
## V8    1674.9
## V9    1675.9
## V10   1734.8
## V11   1836.5
## V12   1755.9
## V13   1836.5
## V14   1727.2
## V15   1744.8
## V16   1713.1
## V17   1655.6
## V18   1620.0
## V19   1569.8
## V20   1542.3
## V21   1496.8
## V22   1426.1
## V23   1378.9
## V24   1379.7
## V25   1433.2
## V26   1475.7
## V27   1544.1
## V28   1693.4
## V29   1260.4
## V30    937.2
```

The following code blocks outline the different statistical analysis I did with the data used. Explanations of the results are described below each figure.

University of Oregon Landfill & Recovered Waste from 1993 to 2021 (Using ggplot2)

```r
plot1 <- ggplot(waste, aes(x = Year)) +
  geom_line(aes(y = All_other, color = "Other"), linewidth=1) +
  geom_line(aes(y = Paper, color = "Paper"), linewidth=1) +
  geom_line(aes(y = Yard_Debris, color = "Yard Debris"), linewidth=1) +
  geom_line(aes(y = Glass_Metal_Plastic, color = "Glass/Metal/Plastic"), linewidth=1) +
  geom_line(aes(y = Food_materials, color = "Food Waste"), linewidth=2) +
  geom_line(aes(y = Landfill, color = "Landfill"), linewidth=1) +
  labs(title = "UO Landfill & Recovered Waste 1993-2021",
       x = "Fiscal Year (July 1st of previous year to June 30th of current year)",
       y = "Weight of Waste in Tons") +
  theme_minimal() +
  scale_color_manual(values = c("Other" = "#d55e00",
                                "Paper" = "#cc79a7",
                                "Yard Debris" = "#0072b2",
                                "Glass/Metal/Plastic" = "#f0e442",
                                "Food Waste" = "#009e73",
                                "Landfill" = "black")) +
  scale_x_continuous(breaks = unique(waste$Year)) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1),
        plot.title = element_text(hjust = 0.5, face = "bold"),
        legend.position = c(0, 1),
        legend.justification = c(0, 1)) +
  labs(color = "Waste Types")
```

```
plot1
```




```
#save plot
#ggsave("LvR.png", plot1, width = 8, height = 6, dpi = 300, bg = "transparent")
```

The color-blind friendly color palette I used for this figure can be found here

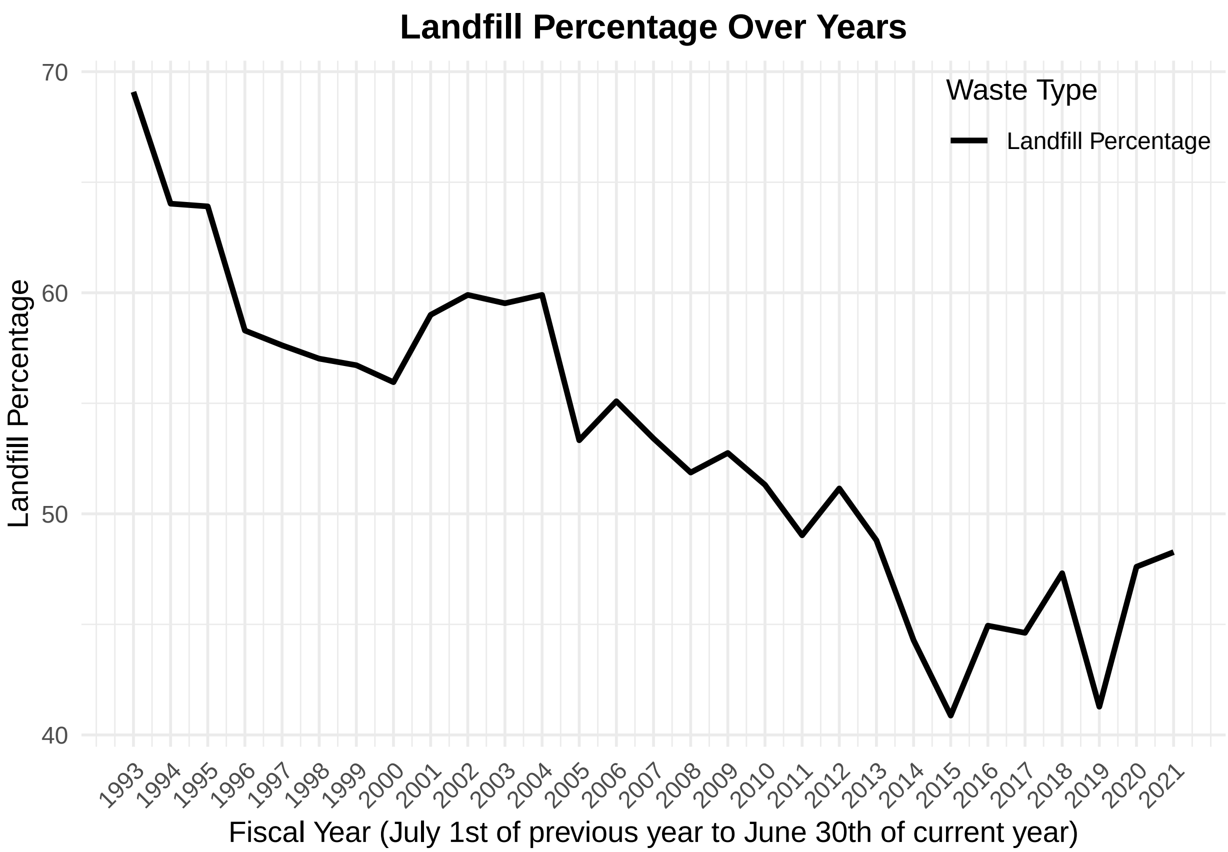
The above plot shows University of Oregon Food Waste that has recovered from Landfill since the Fiscal Year of 1993. We can see that COVID 19 pandemic in 2019-2020 may be the reason that recovered food waste decreased to the lowest levels since 2009.

Landfill Percentage Over the Years

```
waste$Landfill_Percentage <- (waste$Landfill / rowSums(waste[, c("All_other",
                                                                "Paper",
                                                                "Yard_Debris",
                                                                "Glass_Metal_Plastic",
                                                                "Food_materials",
                                                                "Landfill")])) * 100


plot2 <- ggplot(waste, aes(x = Year)) +
  geom_line(aes(y = Landfill_Percentage, color = "Landfill Percentage"), linewidth = 1) +
  labs(title = "Landfill Percentage Over Years",
       x = "Fiscal Year (July 1st of previous year to June 30th of current year)",
       y = "Landfill Percentage") +
  theme_minimal() +
  scale_color_manual(values = c("Landfill Percentage" = "black")) +
```

```r
  scale_x_continuous(breaks = unique(waste$Year)) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1),
        plot.title = element_text(hjust = 0.5, face = "bold"),
        legend.position = c(1, 1),
        legend.justification = c(1, 1)) +
  labs(color = "Waste Type")
plot2
```




```r
#ggsave("LPoY.png", plot2, width = 8, height = 6, dpi = 300, bg = "transparent")
```

While landfill waste at the University of Oregon has **been on the rise** since 2019, again likely due to the pandemic, the linear regression performed below provides clarity on the statistical significance of the overall trend since 1993. As we can see, the linear regression model indicates that there is a significant relationship between 'Year' and 'Landfill_Percentage.' The estimated slope is negative, suggesting **a decrease in 'Landfill_Percentage' over time**. The model has a high R-squared value, indicating a good fit to the data. A summary of the linear model is below.

```r
# in a simple linear regression
lm_result <- lm(Landfill_Percentage ~ Year, data = waste)
summary(lm_result)
```

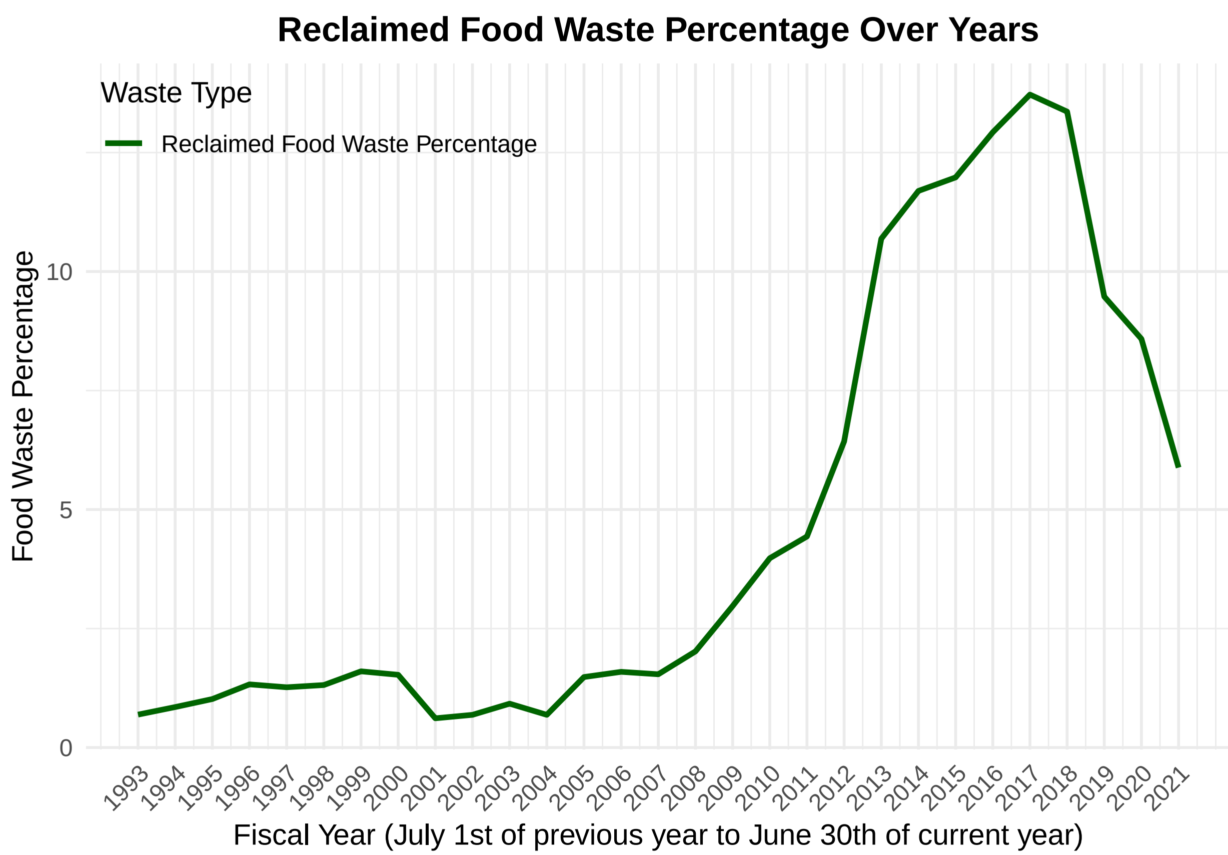
```
##
## Call:
## lm(formula = Landfill_Percentage ~ Year, data = waste)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.4696 -2.6150  0.0664  1.5709  5.4219
```

```
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1557.66337  133.51507   11.67 4.72e-12 ***
## Year          -0.74954    0.06652  -11.27 1.03e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.997 on 27 degrees of freedom
## Multiple R-squared:  0.8246, Adjusted R-squared:  0.8181
## F-statistic: 126.9 on 1 and 27 DF,  p-value: 1.035e-11
```

Reclaimed Food Waste Percentage Over the Years

```
waste$Food_Percentage <- waste$Food_materials / rowSums(waste[, c("All_other",
                                                                  "Paper",
                                                                  "Yard_Debris",
                                                                  "Glass_Metal_Plastic",
                                                                  "Food_materials",
                                                                  "Landfill")]) * 100

plot3 <- ggplot(waste, aes(x = Year)) +
  geom_line(aes(y = Food_Percentage, color = "Reclaimed Food Waste Percentage"), linewidth = 1) +
  labs(title = "Reclaimed Food Waste Percentage Over Years",
       x = "Fiscal Year (July 1st of previous year to June 30th of current year)",
       y = "Food Waste Percentage") +
  theme_minimal() +
  scale_color_manual(values = c("Reclaimed Food Waste Percentage" = "darkgreen")) +
  scale_x_continuous(breaks = unique(waste$Year)) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1),
        plot.title = element_text(hjust = 0.5, face = "bold"),
        legend.position = c(0, 1),
        legend.justification = c(0, 1)) +
  labs(color = "Waste Type")
plot3
```




```r
#save plot
#ggsave("RFWPoY.png", plot3, width = 8, height = 6, dpi = 300, bg = "transparent")
```

Here we see that reclaimed food waste has been **decreasing since 2017**, but the linear regression model performed below clarifies the overall trend of reclaimed food waste which has been on the rise! let's hope it stays that way. The linear model is described below.

```r
# in a simple linear regression
lm_result1 <- lm(Food_Percentage ~ Year, data = waste)
summary(lm_result1)
```

```
##
## Call:
## lm(formula = Food_Percentage ~ Year, data = waste)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.0358 -2.0265 -0.4667  1.9929  4.5867
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -891.72222  121.73105  -7.325 7.03e-08 ***
## Year           0.44663    0.06065   7.364 6.39e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.733 on 27 degrees of freedom
## Multiple R-squared:  0.6676, Adjusted R-squared:  0.6553
```

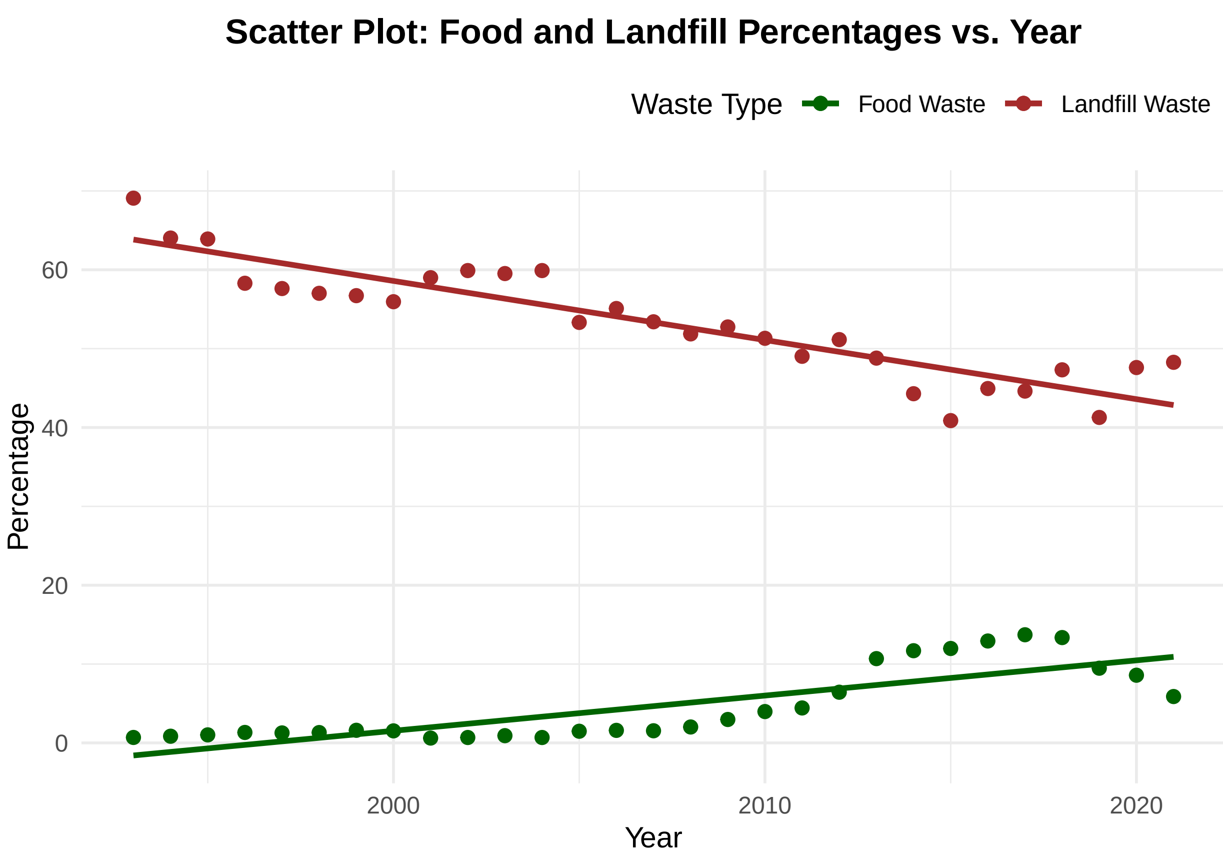
```
## F-statistic: 54.22 on 1 and 27 DF,  p-value: 6.389e-08
```

Scatter Plot: Food and Landfill Percentages vs. Year

```r
plot4 <- ggplot(waste, aes(x = Year)) +
  geom_point(aes(y = Food_Percentage, color = "Food Waste"), size = 2) +
  geom_point(aes(y = Landfill_Percentage, color = "Landfill Waste"), size = 2) +
  geom_smooth(aes(y = Food_Percentage, color = "Food Waste"), method = "lm", se = FALSE) +
  geom_smooth(aes(y = Landfill_Percentage, color = "Landfill Waste"), method = "lm", se = FALSE) +
  labs(title = "Scatter Plot: Food and Landfill Percentages vs. Year",
       x = "Year",
       y = "Percentage") +
  scale_color_manual(values = c("darkgreen", "brown"),
                     name = "Waste Type",
                     labels = c("Food Waste", "Landfill Waste")) +
  theme_minimal() +
  theme(
    plot.title = element_text(hjust = 0.5, face = "bold"),
    legend.position = "top",
    legend.justification = c(1, 0),
    legend.box.just = "right"
  )

plot4
```

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



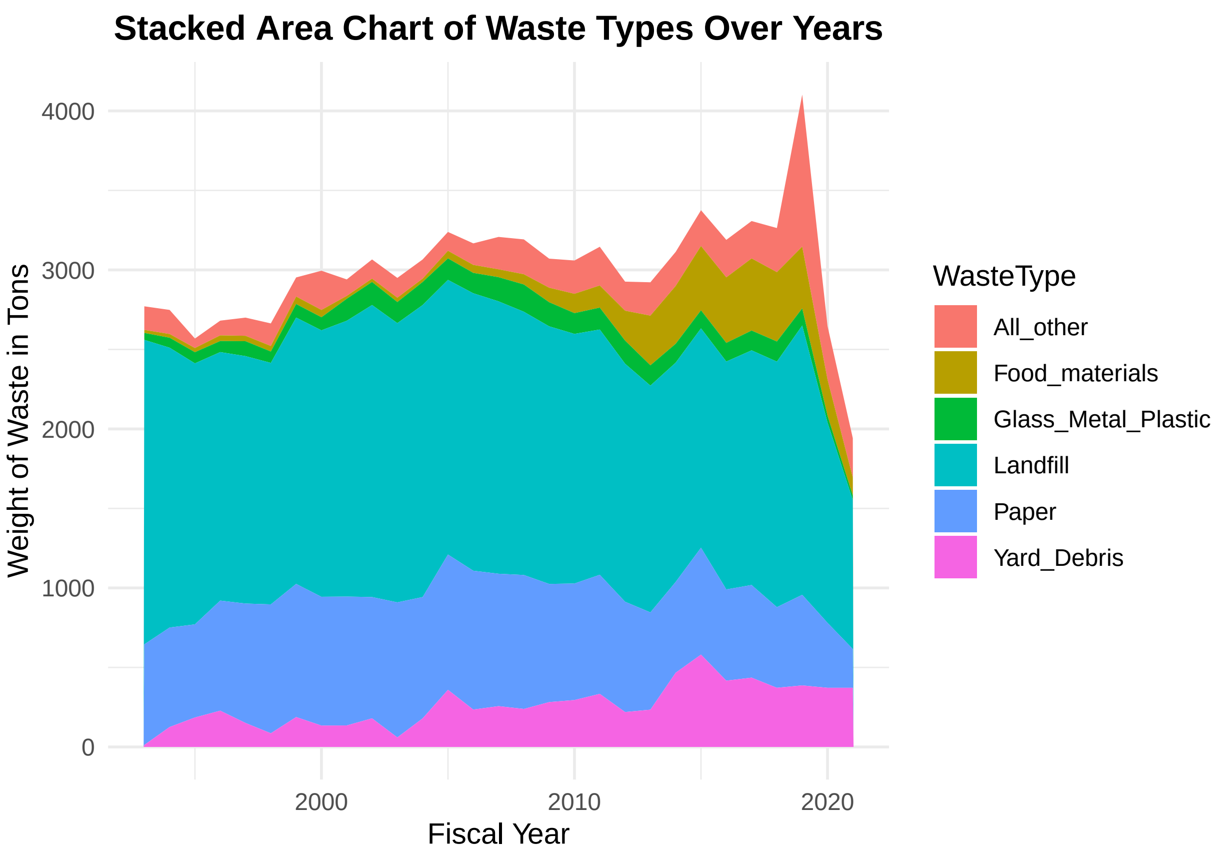
```
#ggsave("FLPY.png", plot4, width = 8, height = 6, dpi = 300, bg = "transparent")
```

In the above plot, we see that as landfill waste decreases, reclaimed food waste increases, indicating that when food waste is not reclaimed, it is included in the regular waste, contributing to the harmful methane gas that is released when food ends up in the landfill. But let's look at this another way.

Stacked Area Chart of Waste Types Over Years

```
waste <- waste[, -c(ncol(waste) - 1, ncol(waste))]
waste_long <- gather(waste, key = "WasteType", value = "Weight", -Year)

# Create the stacked area chart
plot5 <- ggplot(waste_long, aes(x = Year, y = Weight, fill = WasteType)) +
  geom_area() +
  labs(title = "Stacked Area Chart of Waste Types Over Years",
       x = "Fiscal Year",
       y = "Weight of Waste in Tons") +
  theme_minimal()
plot5 <- plot5 + theme(plot.title = element_text(hjust = 0.5, face = "bold"))
plot5
```



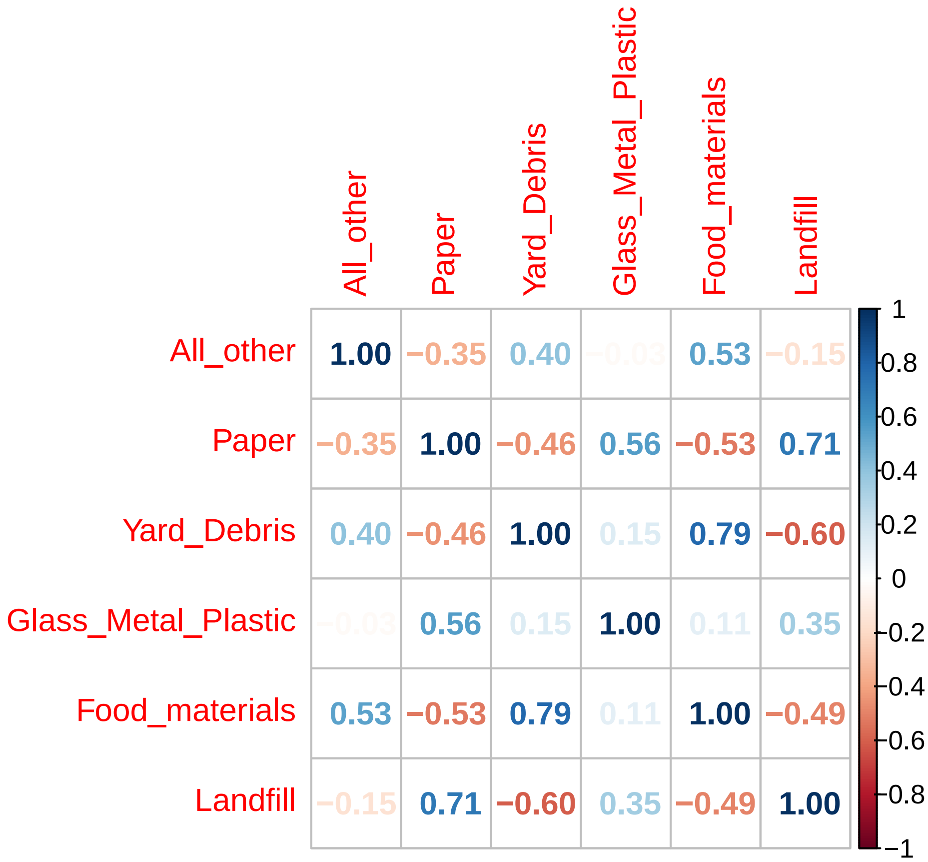

```
#ggsave("WToY.png", plot5, width = 8, height = 6, dpi = 300, bg = "transparent")
```

The above plot gives us a better idea of the total waste produced by the University of Oregon since 1993. we can see clearly that the majority of the waste produced by the UO has been landfill data. As we can see, it looks as though the increase in reclaimed food material is correlated with a decrease in landfill waste, which we will explore later.

Correlation Matrix Between Waste Types

```r
correlation_matrix <- cor(waste[, c("All_other",
                                    "Paper",
                                    "Yard_Debris",
                                    "Glass_Metal_Plastic",
                                    "Food_materials",
                                    "Landfill")])

corr_plot <- corrplot(correlation_matrix, method = "number", addCoef.col = "black")
```




```r
png("correlation_matrix.png", width = 800, height = 600, bg = "transparent")
corrplot(correlation_matrix, method = "number", addCoef.col = "black")
dev.off()
```

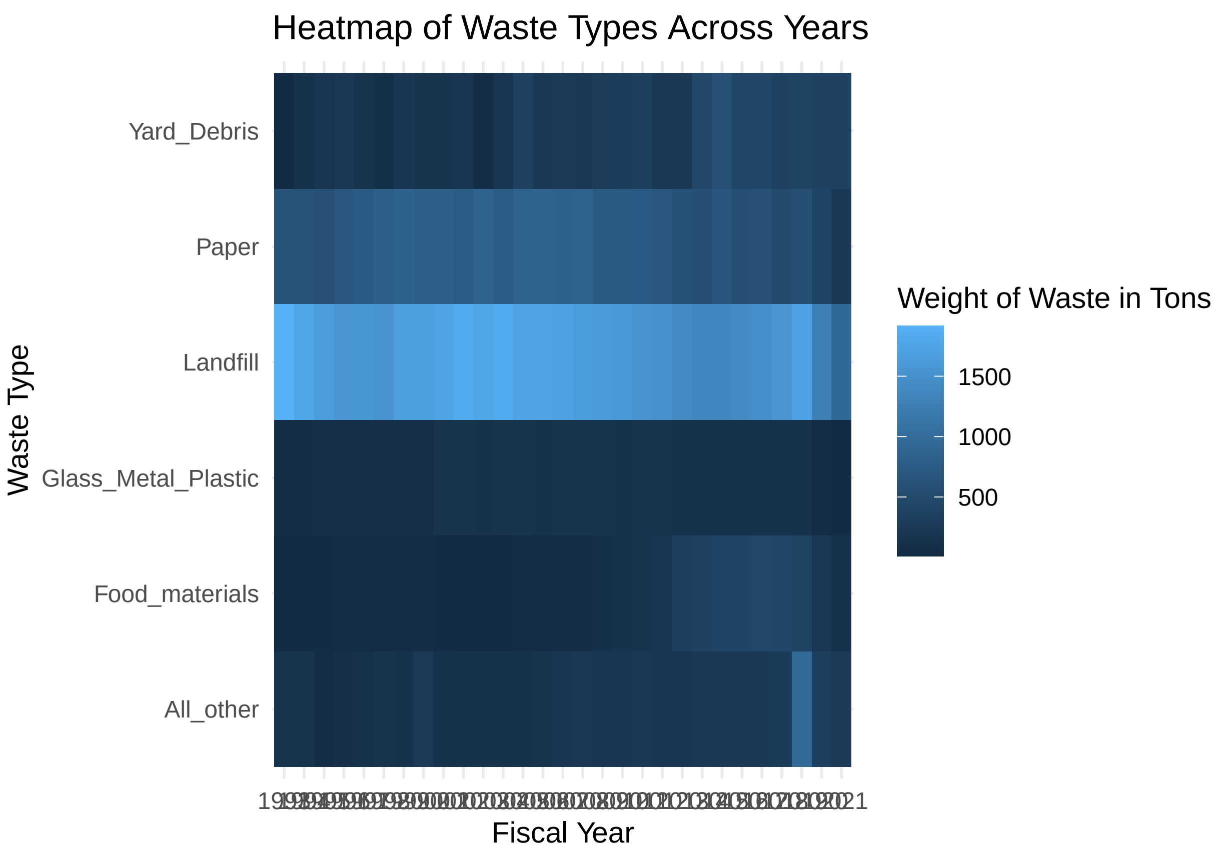
```
## pdf
##   2
```

Here, we can clearly see the negative correlation between food waste and landfill waste. Now we can definitively say that as landfill waste increases, reclaimed food waste decreases, meaning that when food waste is not reclaimed, it is likely ending up in the landfill.

Heatmap of Waste Types Across Years

```r
ggplot(waste_long, aes(x = factor(Year), y = WasteType, fill = Weight)) +
  geom_tile() +
  labs(title = "Heatmap of Waste Types Across Years",
```

```r
       x = "Fiscal Year",
       y = "Waste Type",
       fill = "Weight of Waste in Tons") +
  theme_minimal()
```



We can see that Landfill waste has far surpassed that of other types of waste in terms of weight. Interestingly, we can see that in the years that food waste was higher in weight, landfill is lower! I also noted that reclaimed paper has been on the decline as systems have transitioned to online.

Correlation test of Percentage of Food Waste Vs. Percentage of Landfill Waste

```r
waste$Food_Percentage <- waste$Food_materials / rowSums(waste[, c("All_other",
                                                                   "Paper",
                                                                   "Yard_Debris",
                                                                   "Glass_Metal_Plastic",
                                                                   "Food_materials",
                                                                   "Landfill")]) * 100

waste$Landfill_Percentage <- (waste$Landfill / rowSums(waste[, c("All_other",
                                                                  "Paper",
                                                                  "Yard_Debris",
                                                                  "Glass_Metal_Plastic",
                                                                  "Food_materials",
                                                                  "Landfill")])) * 100

correlation <- cor(waste$Landfill_Percentage, waste$Food_Percentage)
```

```r
# Print the correlation coefficient
cat("Correlation coefficient:", correlation, "\n")
```

```
## Correlation coefficient: -0.8447915
```

```r
# Test for significance (p-value)
cor_test_result <- cor.test(waste$Landfill_Percentage, waste$Food_Percentage)
print(cor_test_result)
```

```
##
##  Pearson's product-moment correlation
##
## data:  waste$Landfill_Percentage and waste$Food_Percentage
## t = -8.2035, df = 27, p-value = 8.266e-09
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.9249220 -0.6927871
## sample estimates:
##        cor
## -0.8447915
```

The negative correlation that we see above serves as a supplemental analysis to the correlation matrix that we saw in the above plot.

Linear model of Food Percentage and Landfill Percentage

```r
model <- lm(Food_Percentage ~ Landfill_Percentage, data = waste)

# Print the summary of the regression model
summary(model)
```

```
##
## Call:
## lm(formula = Food_Percentage ~ Landfill_Percentage, data = waste)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4676 -1.8216 -0.5671  1.9633  5.3234
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)          34.5074     3.6681   9.407 5.18e-10 ***
## Landfill_Percentage  -0.5595     0.0682  -8.204 8.27e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.536 on 27 degrees of freedom
## Multiple R-squared:  0.7137, Adjusted R-squared:  0.7031
## F-statistic:  67.3 on 1 and 27 DF,  p-value: 8.266e-09
```

The above linear model is another way to conceptualize how food percentage is related to landfill percentage.

Below, I found the current number of students on the UO facts page and I used the equation on page 26 in the Lane County Food Waste to Energy Feasibility Study to calculate food waste for this year at UO. In that document, authors source the Massachusetts Department of Environmental Protection (2002).

```r
stustat <- (23202)
# Equation provided in the document:
# 0.35lbs/meal * N students * 405 meals/student/yr

# Food waste generation for this year

fw <- stustat * 405 *.35

# fw is in lbs so /2000 to convert to tons

fwt <- fw/2000

# per student waste in pounds

ps <- fw/stustat
```

Ready for the answer to the joke?

Break it down now y'all!

Thanks for viewing!

Jordan Rodriguez

contact: jrodrig8@uoregon.edu