Dissecting Recycling Dynamics: A Two-Way ANOVA Analysis of Trends in Minnesota's County-Level Waste Management from 1991 to 2017

Introduction

As global awareness of climate change intensifies, various jurisdictions, including states like Minnesota, are adopting waste management strategies to mitigate environmental impacts. This paper focuses on recycling as a pivotal component of these strategies by analyzing a waste management dataset named *waste1data*. The dataset provides comprehensive recycling data from 86 counties in Minnesota, spanning from 1991 to 2017, and includes 49,019 entries across five key variables: Year, County, Category, Res Tons (Residential Tons), and CII Tons (Commercial/Institutional/Industrial Tons).

The Year variable is numerical and discrete, referring to the exact year in which the data was collected. The County is a categorical and nominal variable, on the name of the county in Minnesota where the recycling data was collected. The Category variable is also categorical and nominal; it refers to the type of recycling material, which is categorized into seven levels: Paper, Metal, Glass, Plastic, Hazardous, Other, and Organic. These seven levels of Category variables assist in the analysis of recycling distribution by material type. The Res Tons and CII Tons variables are both continuous and numerical, referring to the weight of recycling collected from residential as well as commercial, institutional, and industrial sources, respectively.

This study is guided by two main research questions. The first question seeks to understand whether there is a significant difference in the mean weight of waste recycled between two sources: residential areas and commercial, industrial, and institutional sources. To explore this, we transform the Residential Tons (Res Tons) and Commercial/Institutional/Industrial Tons (CII Tons) variables from wide to long format, creating two new variables: Type and Recycle Tons. The Type variable, which is categorical and nominal, identifies the recycling source as either residential or commercial/institutional/industrial and serves as an independent variable. Recycle Tons, a continuous numerical variable represents the weight of recycling collected from two sources (Res and CII) and acts as a dependent variable.

The second question seeks to determine if significant differences exist in the mean weight of waste recycled across different material categories, such as Paper, Metal, Glass, Plastic, Hazardous, Other, and Organic. For this analysis, we use the Category variable as an independent variable and the Recycle Tons variable as the dependent variable to assess variations in recycling amounts by material categories.

Together, these questions aim to uncover specific patterns that could lead to improvements in recycling methods, depending on both the material categories and the type of materials recycled.

Null and Alternative Hypotheses

- **Main Effect of Category**
  - **Null Hypothesis (H_0):** The means across all categories are equal.
    - H_0: $\mu_{Paper} = \mu_{Metal} = \mu_{Glass} = \mu_{Plastic} = \mu_{Hazardous} = \mu_{Other} = \mu_{Organic}$
  - **Alternative Hypothesis (H_a):** Not all category means are equal.
    - H_a: At least two means are different.
- **Main Effect of Type**
  - **Null Hypothesis (H_0):** The mean weight of recycled tons is the same for residential and commercial/industrial/institutional (CII) collections.
    - H_0: $\mu_{Res} = \mu_{CII}$
  - **Alternative Hypothesis (H_a):** The mean weight of recycled tons for residential collection is not equal to that for commercial/industrial/institutional (CII).
    - H_a: $\mu_{Res} \neq \mu_{CII}$
- **Interaction Between Type and Category**
  - H_0: There is no interaction between Type and Category
  - H_a**:** There is interaction between Type and Category

Methods

Before conducting the statistical analysis, preliminary steps such as data exploration, visualization, and transformation are essential to ensure the quality and integrity of the dataset. These initial measures include checking for and addressing any missing values, outliers, or anomalies that could affect the outcomes. Importantly, we transform data from wide to long format to better suit the analysis needs.

For this project, we employ a two-way ANOVA to analyze the dataset, a choice driven by some compelling reasons that align with our research objectives. Specifically, our main goal is to compare mean differences across two independent factors: Type and Category. This makes two-way ANOVA the most appropriate statistical test for our purposes.

Furthermore, two-way ANOVA does not only examine the individual effects of the source type (Residential vs. Commercial/Institutional/Industrial) and recycling category (Paper, Metal, Glass, Plastic, Hazardous, Other, Organic) on the amount of waste recycled, but it also allows for the analysis of interaction effects between these factors. This capability is crucial for understanding whether the influence of one factor depends on the level of the other. It provides a deeper insight into the recycling behaviors across different materials and source types, offering a nuanced view that is vital for effective recycling strategy development.

Conclusion

The test statistics and p-values provided by the two-way ANOVA offer informative insights for the Category, Type, and their interaction; and they help us answer our two main research questions. Importantly, the p-values for each variable and their interaction are highly significant, well below the conventional alpha threshold of 0.05. Specifically, as shown in Figure 1, the F-value for the Category variable is 342.1, and the p-value is less than 2e-16, significantly below 0.05. This result leads us to reject the null hypothesis H_0 for the Category, strongly suggesting a significant main effect of the category on the amount of recycling. This indicates that the mean weights of recycled tons vary significantly among the different categories (Paper, Metal, Glass, Plastic, Hazardous, Other, and Organic). Consequently, this helps to answer the first main research question which proves that there are indicates significant differences exist in the mean weight of waste recycled across different material categories, such as Paper, Metal, Glass, Plastic, Hazardous, Other, and Organic, as confirmed by Figures 2, with that at least two means of the weight of Recycle Tons of are statistically significantly different when comparing different levels within Categories.

In terms of the Type variable, which distinguishes between Residential and Commercial/Industrial/Institutional (CII) sources, the reported F value is 143.4 with a p-value less than 2e-16, again below 0.05. This result leads us to reject the null hypothesis H_0 for Type. This gives us evidence that the mean weights of recycled tons significantly differ between Residential and CII sources. As a result, this helps us to answer the second research question which prove that this a significant difference in the mean weight of waste recycled between two sources: residential areas and commercial, industrial, and institutional sources as confirmed by Figure 3.

Additionally, the interaction between Category and Type also shows significant statistical values, with an F value of 109.3 and a p-value less than 2e-16, which is less than 0.05. We reject the null hypothesis H_0 for no interaction effect between Category and Type. This indicates there is significant evidence that there is a significant interaction between Category and Type. This confirms that the relationship between Category and Type significantly influences recycling behaviors and that these factors do not operate independently.

```
                 Df    Sum Sq   Mean Sq F value Pr(>F)    
Category          6 1.463e+10 2.438e+09   342.1 <2e-16 ***
Type              1 1.022e+09 1.022e+09   143.4 <2e-16 ***
Category:Type     6 4.676e+09 7.793e+08   109.3 <2e-16 ***
Residuals     98024 6.986e+11 7.127e+06                   
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 1: Output from Two-Way ANOVA

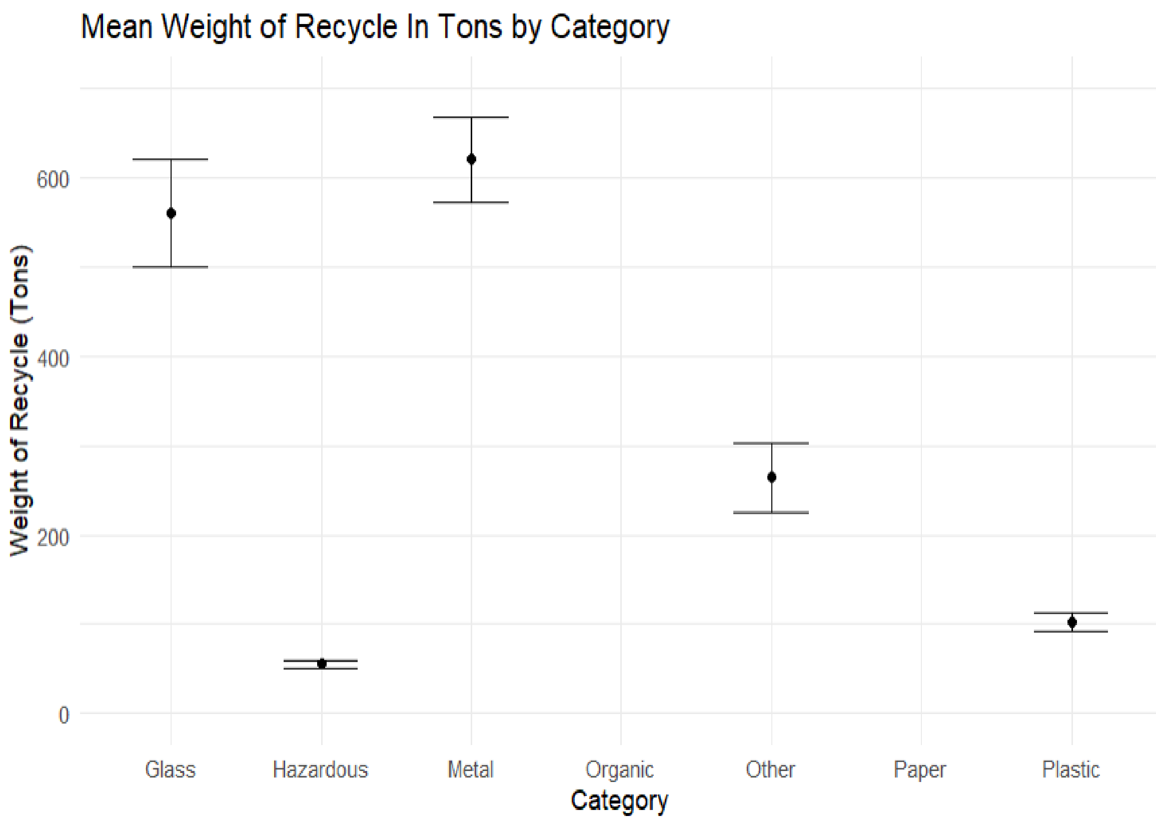

Figure 2: Mean Weight of Recycle in Tons by Category

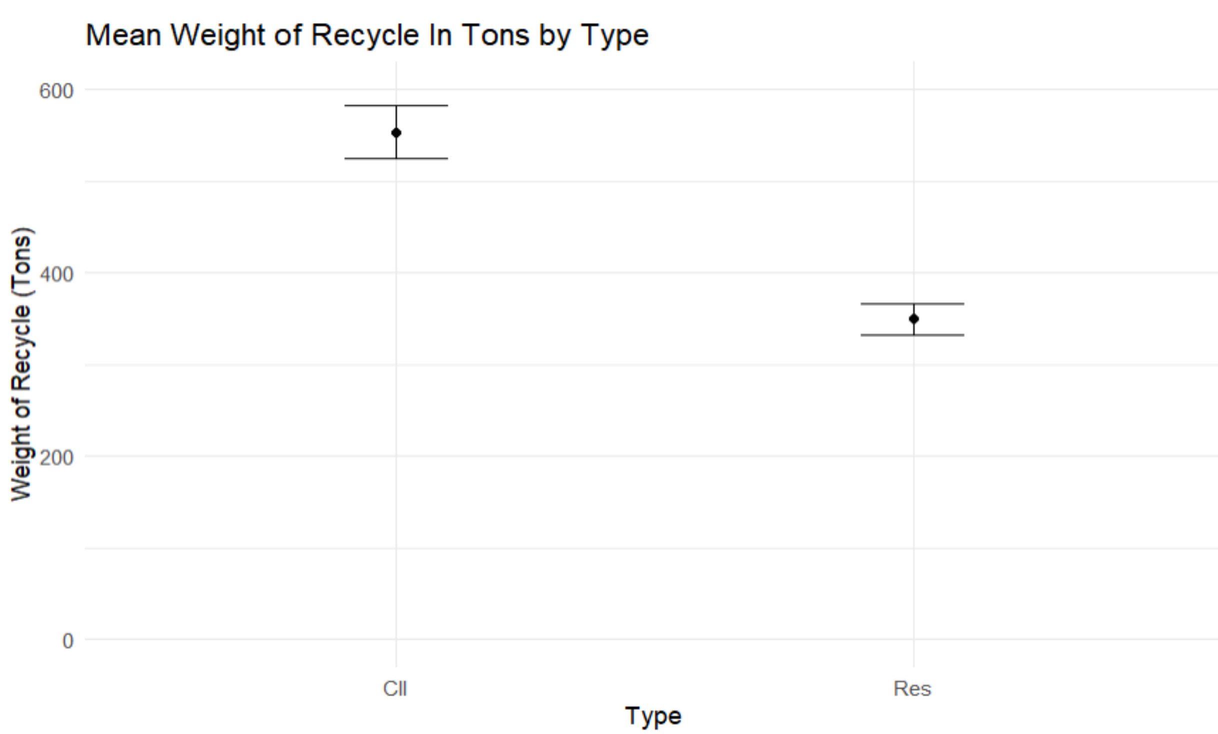

Figure 3: Mean Recycle in Tons by Type

Perhaps the most interesting finding is the significance of the interaction effect, which could indicate complex dynamics between the source of the waste (Res and CII) and the Category of material (Paper, Metal, Glass, Plastic, Hazardous, Other, Organic) being recycled. This suggests that recycling programs might need to be focused specifically on different categories of recycling (Paper, Metal, Glass, Plastic, Hazardous, Other, Organic) and their Type of source (Res or CII) to maximize efficiency.

To gain a better understanding of the interaction effects between the Category and Type variables within the dataset, a more focused analytical approach was adopted. The data were filtered to retain only one Type at a time, and individual one-way ANOVAs were conducted for each, specifically for Residential (Res) and Commercial, Institutional, and Industrial (CII). This method provides a clearer analysis of the differences within material categories under each type setting.

Initially, the data were filtered to include only the Residential type. A one-way ANOVA was then performed to test for significant differences among the material categories within residential settings. Based on Figure 4, the Category variable has an F-value of 134.2, and the p-value is less than 2e-16, which is significantly below 0.05. This indicates that at least two mean weights of recycled tons among the seven categories significantly differ within the Residential type. Tukey's multiple comparisons of means revealed significant differences among several pairs of categories within the residential type. However, no significant differences were observed between Paper and Glass, Plastic and Hazardous, or Organic and Metal within the residential type. These findings suggest variable recycling rates for different materials within residential settings but indicate similar treatment for certain materials.

```
             Df    Sum Sq   Mean Sq F value Pr(>F)    
Category      6 2.840e+09 473278697   134.2 <2e-16 ***
Residuals 49012 1.728e+11   3525725                   
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 4: one-way ANOVA of Category Within Residential Type

Similarly, the dataset was filtered to include only entries classified under CII (Commercial, Institutional, and Industrial). A one-way ANOVA was conducted to test for significant differences among the material categories (Paper, Metal, Glass, Plastic, Hazardous, Other, Organic) within CII settings. According to Figure 5, the Category variable has an F-value of 255.8, and the p-value is less than 2e-16, well below 0.05. This suggests that at least two mean weights of recycled tons among the seven levels of Category significantly differ within the CII type. Tukey's multiple comparisons of means indicate significant differences among most pairs; however, no significant differences were observed between Other and Glass, and between Plastic and Hazardous in the CII type. This pattern highlights diverse recycling behavior in CII settings, with some materials exhibiting comparable recycling weights.

```
                 Df    Sum Sq   Mean Sq F value Pr(>F)    
Category          6 1.646e+10 2.744e+09   255.8 <2e-16 ***
Residuals     49012 5.258e+11 1.073e+07                   
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 5: one-way ANOVA of Category Within Commercial, Institutional, and Industrial (CII) Type

It is important to note some limitations of this analysis. The ANOVA assumes homogeneity of variances, and if this condition is not met, the validity of the results could be compromised. Additionally, the analysis does not account for potential confounding variables that were not included in the model but might still influence the amount of waste recycled. Lastly, the dataset is limited to data from Minnesota, which may not be generalizable to other regions or states.