Data Appendix to "Smith Compost Analysis"

Kate Ginder

Contents

1 Appendix description

This Data Appendix documents the data used in "Smith Compost Analysis". It was prepared in a Rmarkdown document that contains both the documentation and the R code used to prepare the data used in the final estimation. It also includes descriptive statistics for both the original data and the final dataset, with a discussion of any issues of note. This data is a time series and will record pounds of compost in a college dining hall.

The datasets used directly by the final analysis are saved in `processed-data/` at the end of this file.

2 Raw data

This section documents the data sets used in this analysis.

2.1 Dataset description

Citation: Smith College Dining Services

Date Downloaded: 04/17/2020

Filename: Compost Tracker 3.0.xlsx.

Unit of observation: amount of compost recorded in dining halls daily

Dates covered: February 2020 - April 2020

2.1.1 To obtain a copy

To obtain a copy of this data set please contact Susan Sayre at ssayre@smith.edu. Due to the COVID-19 pandemic the spring semester 2020 was moved to remote learning and the dining halls were closed. The data analysed below is artificially created. Numbers are based upon one dining hall over the course of a week, but do not represent the impact of the study.

2.1.2 Importable version

Filename: importable-data/Raw Data Seminar Paper/Compost Tracker 6.0.csv

The following changes were made to create the importable files.

1. The file was originally opened in excel on a Mac.

2. The header reading "Composting Feb & April" was deleted. It was causing the variable names to import incorrectly.

3. Variable names were edited to allow R to read them.

4. The document was then saved as a csv file.

3 Variable descriptions

The following data is from King Scales and Cutter Ziskind dining halls at Smith College.

- **dates:** Date of the month.

- **Dayoftheweek:** Day of the week the meal is served on.
- **#ofplatesking:** Number of plates per night used in King dining hall.
- **lbcompostingking:** Pounds of compost per night used in King dining hall.
- **#ofplatescutter:** Number of plates per night used in Cutter dining hall.
- **lbcompostingcutter:** Pounds of compost per night used in Cutter dining hall.
- **Meal_number:** Rotating menu cycle that coordinates to a different number.

3.0.1 Data import code and summary

```
library(readr)
importable_data <- read_csv("/Users/Kate/Desktop/seminar-paper-kginder/Raw Data Seminar Paper/importable
```

```
## Parsed with column specification:
## cols(
##   dates = col_character(),
##   Dayoftheweek = col_character(),
##   `#ofplatesKing` = col_double(),
##   lbcompostKing = col_double(),
##   `#ofplatescutter` = col_double(),
##   lbcompostCutter = col_double(),
##   `meal number` = col_double()
## )
```

```
#use getwd() get current directory


Compost_data <-importable_data %>%
  #rename the variables
  rename(king_plates = `#ofplatesKing`,
         king_compost = lbcompostKing,
         cutter_plates = `#ofplatescutter`,
         cutter_compost = lbcompostCutter)  %>%
  #create new variable house
pivot_longer(contains("_"), names_to = c("house","variable"), names_sep = '_', values_to = "values") %>%
  pivot_wider(names_from = "variable", values_from = "values") %>%
  #edit data so date is read as a numerical date
  mutate(date_var = as.Date(dates, "%m/%d/%y")) %>%
  #create new variable cycle
  mutate(cycle = case_when(date_var <= as.Date("2020-03-01") ~ 1,
date_var <= as.Date("2020-04-04") ~ 2,
date_var <= as.Date("2020-05-03") ~ 3))
```

```
#create a compost per plate variable
Compost_data <- Compost_data %>%
 mutate(compost_plate = compost/plates)
```

```
#create a dummy variable for poster date
Compost_data <- Compost_data %>%
  mutate(poster_date = case_when(house == "cutter" ~ 0,
date_var <= as.Date("2020-04-04") ~ 0,
date_var <= as.Date("2020-05-2") ~ 1))
```

4 Analysis Variables

This section includes a description of all the variables that are used in the final analysis. At the end of this section, these variables are saved in the processed_data folder of the repository.

4.1 Variables used in the final analysis

- **date_var:** Date of the month read as a numerical value.

- **house:** House variable, King or Cutter.
- **plates:** Number of plates counted per meal.

- **compost:** Pounds of compost collected per night.
- **cycle:** Classifies the rotating menu into three distinct cycles.Cycle 1: 2/2-2/29, Cycle 2: 3/1-4/4, Cycle 3, 4/5-5/2
- **poster_date:** Dummy variable for posters (poster date 0: 2/2 - 2/29) (posters poster date 1: 4/5 - 5/2).
- **compost_plate** The amount of compost in (lb) divided by the number of plates.

The variable 'date_var' originally came from the variable 'dates.' 'Dates' was read as a categorical variable, and it needed to be recognized as a numerical number. These dates space the length of the experiment. The variable 'house' was created by using the pivot function. The data was originally set up for compost and plates in King house and compost and plates in Cutter. By using the pivot() function 'house' was able to become its own variable with the option for the two different houses. 'Plates' record the number of plates used per meal. This figure is used to stand as a proxy for the number of students eating in the dining hall. 'Compost' is recorded in pounds after each mean. 'Cycle' is derived from the 'meal number' variable, from 1-28 indicating the rotating meals for each cycle of the Smith College menu cycle. There are three separate menu cycles. 'Poster_date' is a dummy variable representing the categorical variable the study is looking at. When the posters were not up, the variable was 0. When the posters were up, the variable was 1. The variable 'compost_plate' is the total amount of compost per night in pounds divided by the number of plates used. This will allow us to look at the amount of compost left on each plate per night, and show a more accurate description to if there is a change due to the intervention.

5 Summary Statistics

```
summary(Compost_data)
```

```
dates           Dayoftheweek        meal number       house
```

Length:168 Length:168 Min. : 1.00 Length:168
Class :character Class :character 1st Qu.: 7.75 Class :character
Mode :character Mode :character Median :14.50 Mode :character
Mean :14.50
3rd Qu.:21.25
Max. :28.00
plates compost date_var cycle
Min. :121.0 Min. :14.00 Min. :2020-02-02 Min. :1.000
1st Qu.:215.5 1st Qu.:31.44 1st Qu.:2020-02-22 1st Qu.:1.000
Median :334.0 Median :43.38 Median :2020-03-21 Median :2.000
Mean :288.3 Mean :42.54 Mean :2020-03-18 Mean :1.988
3rd Qu.:349.2 3rd Qu.:53.50 3rd Qu.:2020-04-11 3rd Qu.:3.000
Max. :371.0 Max. :78.75 Max. :2020-05-02 Max. :3.000
compost_plate poster_date
Min. :0.09503 Min. :0.0000
1st Qu.:0.13014 1st Qu.:0.0000
Median :0.14635 Median :0.0000
Mean :0.14725 Mean :0.1667
3rd Qu.:0.16259 3rd Qu.:0.0000
Max. :0.21694 Max. :1.0000

```
#new data frame for king data
king_data <- Compost_data %>% filter(house == "king")
```

```
stargazer::stargazer(as.data.frame(filter(king_data, cycle == 1)),
                title = "King Summary Statistics Cycle 1",
                label = "fullResults",
                header = F,
                summary = T,
                type = "latex",
                keep = c("compost", "plates", "compost_plate"))
```

Table 1: King Summary Statistics Cycle 1

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
plates	29	284.966	79.403	121	210	345	370
compost	29	41.698	12.508	17.250	32.000	52.250	64.250
compost_plate	29	0.147	0.021	0.109	0.129	0.162	0.188

```
stargazer::stargazer(as.data.frame(filter(king_data, cycle == 3)),
                title = "King Summary Statistics Cycle 3",
                label = "fullResults",
                header = F,
                summary = T,
                type = "latex",
                keep = c("compost", "plates", "compost_plate"))
```

Table 2: King Summary Statistics Cycle 3

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
plates	28	291.679	78.053	138	230.2	351	368
compost	28	40.723	14.245	14	29.1	50.8	71
compost_plate	28	0.139	0.026	0.095	0.119	0.152	0.196

```
#skim(select(filter(king_data, cycle == 1), c("compost", "plates", "compost_plate")))
```

```
sd(filter(king_data, cycle == 1 )$plates)
```

[1] 79.4029

```
sd(filter(king_data, cycle == 3 )$plates)
```

[1] 78.05271

```
sd(filter(king_data, cycle == 1 )$compost)
```

[1] 12.50828

```
sd(filter(king_data, cycle == 3 )$compost)
```

[1] 14.24477

```
sd(filter(king_data, cycle == 1 )$compost_plate)
```

[1] 0.02086918

```
sd(filter(king_data, cycle == 3 )$compost_plate)
```

[1] 0.02598106

```
#new data frame for cutter data
cutter_data <- Compost_data %>% filter(house == "cutter")


stargazer::stargazer(as.data.frame(filter(cutter_data, cycle == 1)),
                title = "Cutter Summary Statistics Cycle 1",
                label = "fullResults",
                header = F,
                summary = T,
                type = "latex",
                keep = c("compost", "plates", "compost_plate"))
```

Table 3: Cutter Summary Statistics Cycle 1

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
plates	29	289.000	77.971	133	208	353	362
compost	29	41.741	13.887	17	29.2	53.5	61
compost_plate	29	0.143	0.021	0.106	0.130	0.159	0.198

```
stargazer::stargazer(as.data.frame(filter(cutter_data, cycle == 3)),
                title = "Cutter Summary Statistics Cycle 3",
                label = "fullResults",
                header = F,
                summary = T,
                type = "latex",
                keep = c("compost", "plates", "compost_plate"))
```

Table 4: Cutter Summary Statistics Cycle 3

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
plates	28	289.214	78.704	134	224.2	348.8	364
compost	28	43.652	12.261	21.750	35.500	53.500	59.750
compost_plate	28	0.152	0.017	0.114	0.141	0.164	0.182

```
sd(filter(cutter_data, cycle == 1)$plates)
```

[1] 77.97069

```
sd(filter(cutter_data, cycle == 3)$plates)
```

[1] 78.70449

```
sd(filter(cutter_data, cycle == 1)$compost)
```

[1] 13.88706

```
sd(filter(cutter_data, cycle == 3)$compost)
```

[1] 12.26111

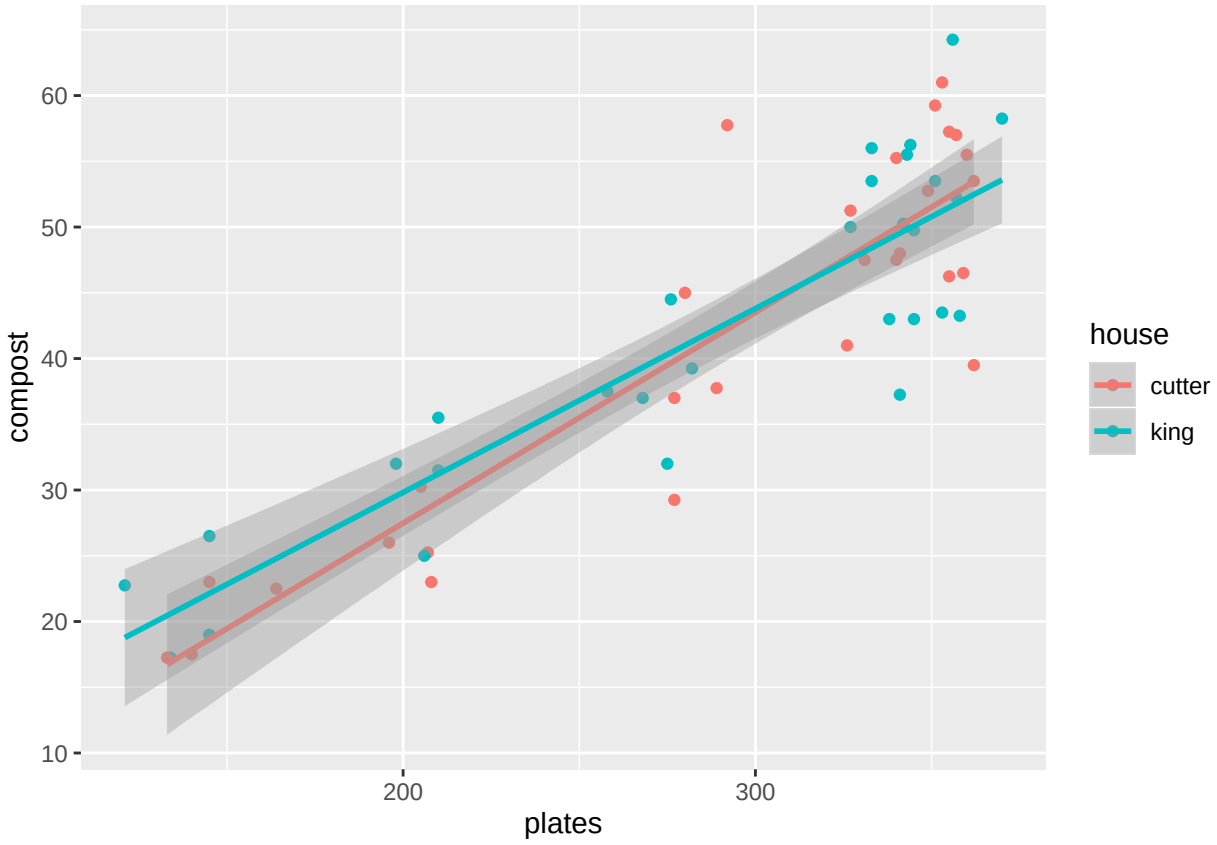
```
sd(filter(cutter_data, cycle == 1 )$compost_plate)
```

[1] 0.02079428

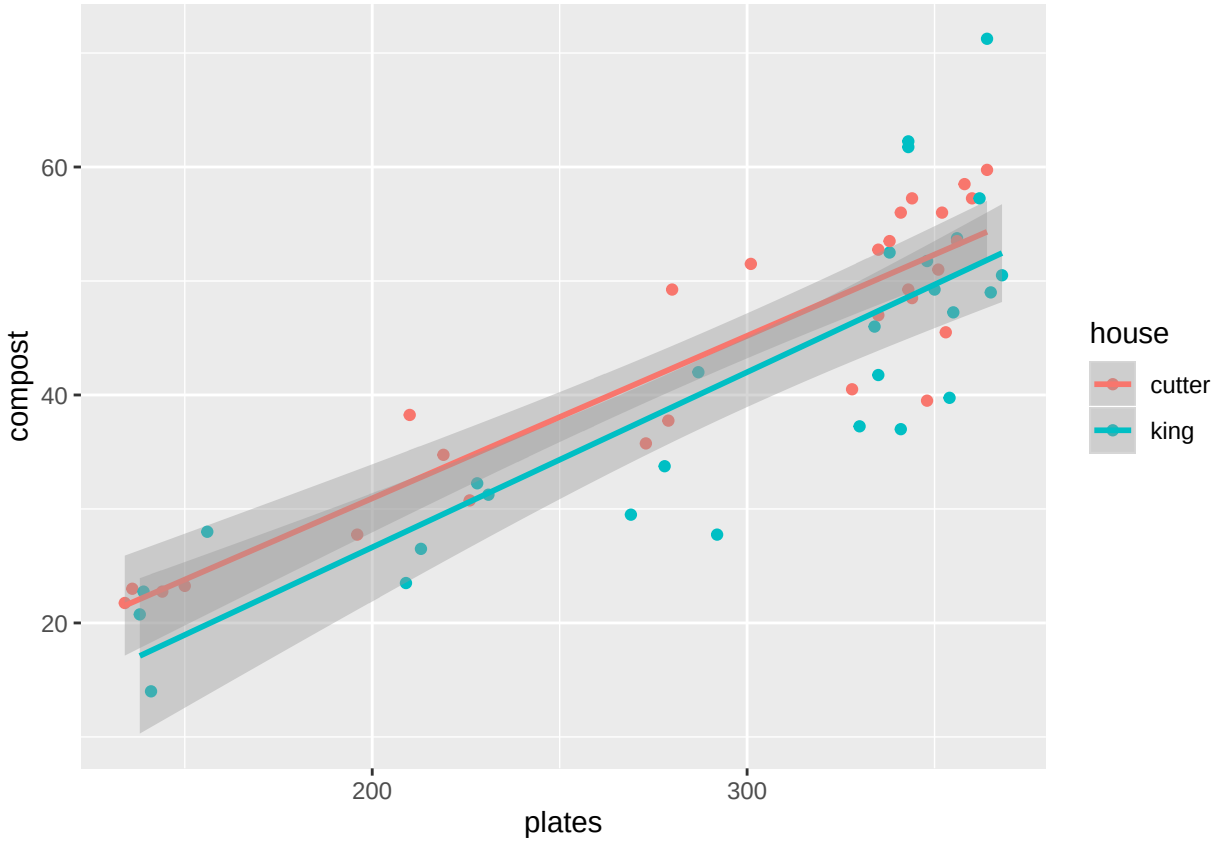
```
sd(filter(cutter_data, cycle == 3 )$compost_plate)
```

[1] 0.01658584

```
ggplot(data = filter(Compost_data, cycle == 1), aes(plates, compost, color = house)) + geom_point() + g
```



```
ggplot(data = filter(Compost_data, cycle == 3), aes(plates, compost, color = house)) + geom_point() + g
```

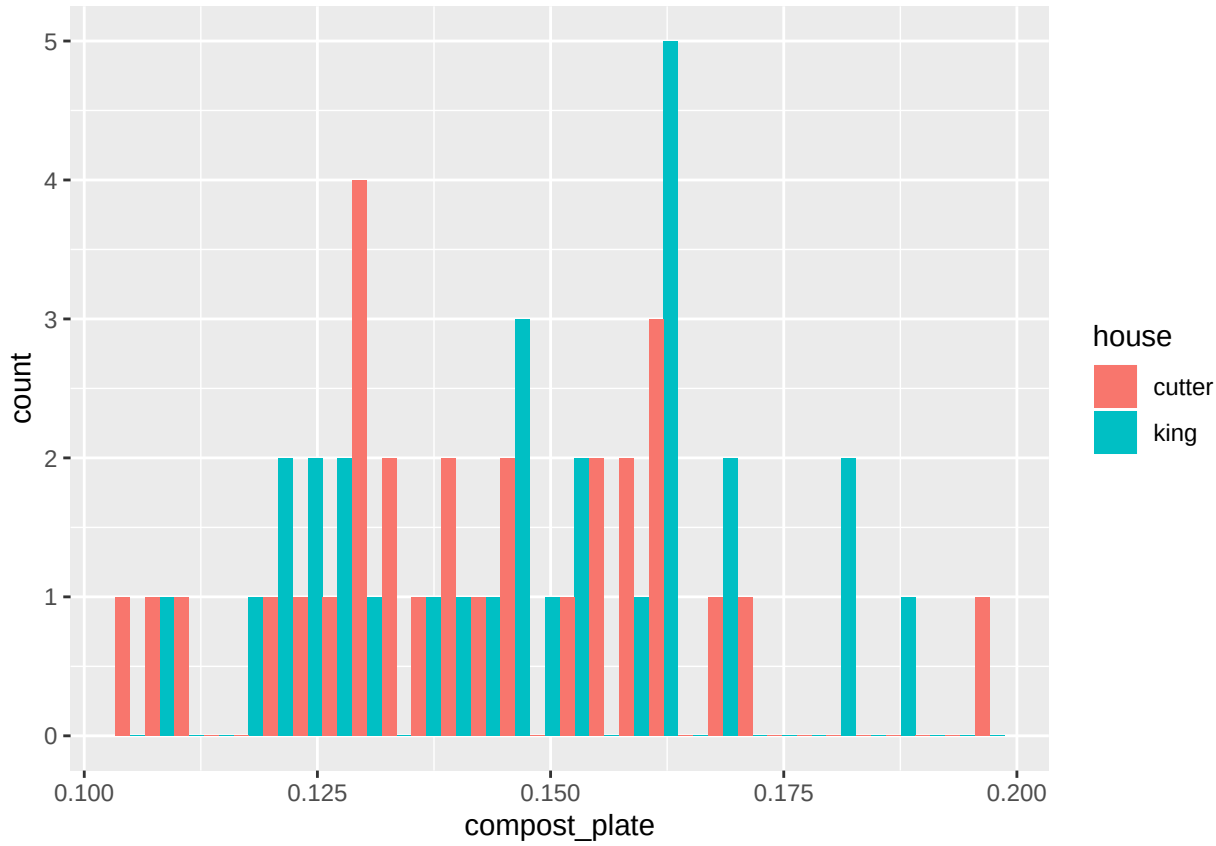


6 Hisograms

```
cycle1_results <- Compost_data %>%
  filter(cycle == 1)

ggplot(data = cycle1_results, aes(x = compost_plate, fill = house)) + geom_histogram(position = "dodge"
```

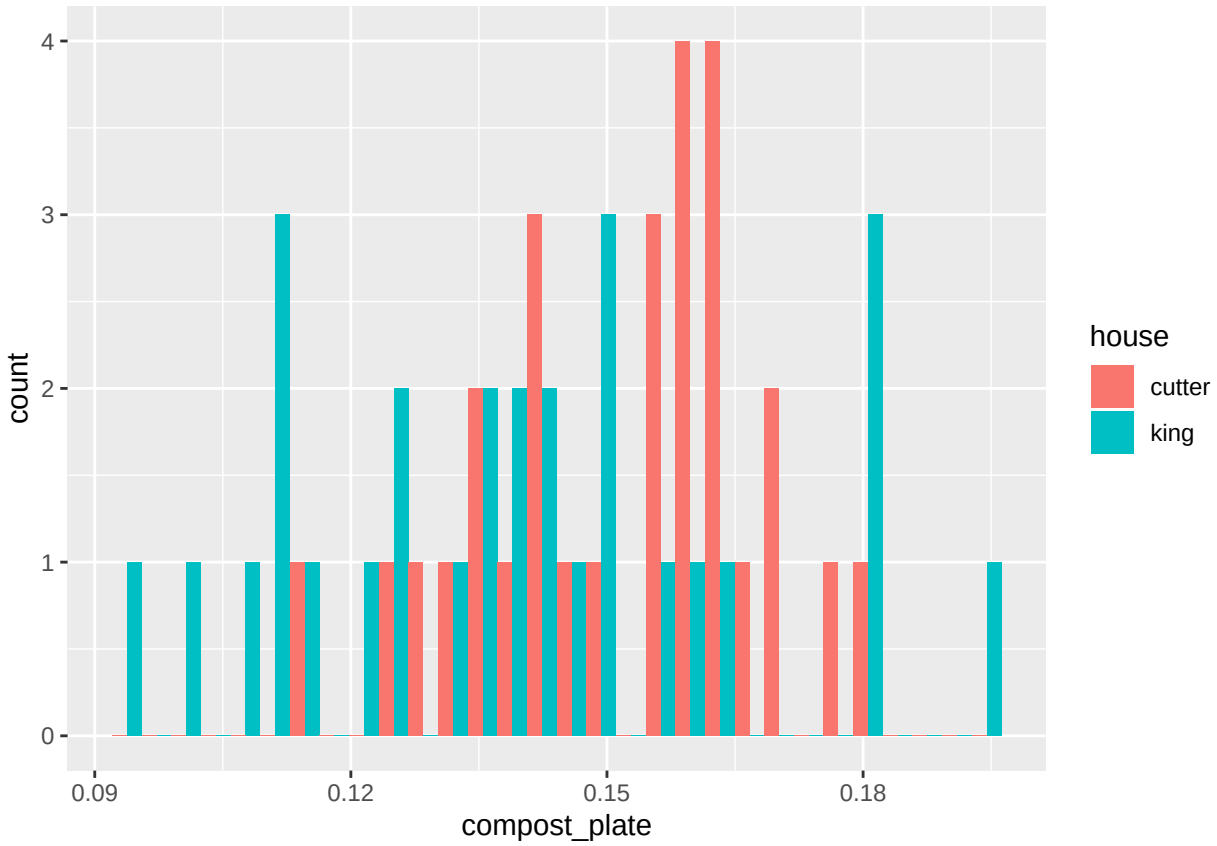
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
cycle3_results <- Compost_data %>%
  filter(cycle == 3)

ggplot(data = cycle3_results, aes(x = compost_plate, fill = house)) + geom_histogram(position = "dodge"
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



7 Correlation

```
cor(king_data$compost, king_data$plates)
```

[1] 0.8345402

```
cor(cutter_data$compost, cutter_data$plates)
```

[1] 0.884931

8 Regression

```
compostmod1 = lm(compost_plate ~ poster_date + house, data = Compost_data)
summary(compostmod1)
```

Call: lm(formula = compost_plate ~ poster_date + house, data = Compost_data)

Residuals: Min 1Q Median 3Q Max -0.045193 -0.017765 -0.000504 0.015011 0.069522

Coefficients: Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.147421 0.002566 57.461 <2e-16 ** *poster_date -0.012200 0.005442 -2.242 0.0263*

houseking 0.003728 0.004057 0.919 0.3594
— Signif. codes: 0 ‘’ ***0.001*** ’’ *0.01* ’’ 0.05 ‘.’ 0.1 ’ ’ 1

Residual standard error: 0.02351 on 165 degrees of freedom Multiple R-squared: 0.0296, Adjusted R-squared: 0.01784 F-statistic: 2.517 on 2 and 165 DF, p-value: 0.08382

```
linearMod = lm(compost_plate ~ poster_date + house, data = filter(Compost_data, cycle != 2))
summary(linearMod)
```

Call: lm(formula = compost_plate ~ poster_date + house, data = filter(Compost_data, cycle != 2))

Residuals: Min 1Q Median 3Q Max -0.043915 -0.016657 -0.000104 0.013903 0.056793

Coefficients: Estimate Std. Error t value Pr(>|t|)
(Intercept) 1.474e-01 2.842e-03 51.847 <2e-16 *** poster_date -8.388e-03 5.685e-03 -1.475 0.143
houseking -2.502e-05 4.895e-03 -0.005 0.996
— Signif. codes: 0 ‘’ ***0.001*** ’’ *0.01* ’’ 0.05 ‘.’ 0.1 ’ ’ 1

Residual standard error: 0.02146 on 111 degrees of freedom Multiple R-squared: 0.02836, Adjusted R-squared: 0.01086 F-statistic: 1.62 on 2 and 111 DF, p-value: 0.2025

Table 5: Regession Results

	Dependent variable:
	compost_plate
poster_date	−0.008
	(0.006)
houseking	−0.00003
	(0.005)
Observations	114
R^2	0.028
Adjusted R^2	0.011
Residual Std. Error	0.021 (df = 111)
F Statistic	1.620 (df = 2; 111)
Note:	*p<0.1; **p<0.05; ***p<0.01

9 Discussion of Data

By first looking at the summary statistics for Cutter and King, one can see the data is very similar. Because of the size of the two houses, this is to be expected. During the baseline period, Cycle 1, King and Cutter have very similar mean pounds of compost: 41.70 and 41.74, respectively. They also have similar plate counts: 285 and 289, respectively. When the trial begins, Cycle 3, those numbers shift slightly. When King begins the treatment, the mean pounds of compost drops to 40.72. Cutter, with no trial implemented, has a mean compost of 43.65 pounds. The mean plates in King during the trial increases slightly to 291.7 while the plates in Cutter moves somewhat to 289.2. Looking more closely at the change in compost per plate in Cycle, 1 King has a mean of .147 pounds of compost per plate. In Cycle 3 King's compost per plate fall slightly to .138 pounds. This means that for each plate used, there is slightly less compost on it. In Cutter the mean compost per plate in Cycle 1 is .1429 pounds. In Cycle 3 this increases slightly to .1519 pounds of compost per plate.

The standard deviation for King compost in the baseline period (Cycle 1) is 12.5 lbs but increases slightly to 15.48 lbs during the experiment (Cycle 3). Overall, there is minimal spread for total amount of compost each night. The difference between the 1st and 3rd quartile is at most 20 pounds. Cutter's standard deviation is similar with a spread of 13.88 lbs during the baseline (Cycle 1) and 16.10 lbs during the trial (Cycle 3). For either period, the largest difference between the 1st and 3rd quartile is 30 pounds. In King's Cycle 1 the standard deviation for compost per plate is .02. In Cycle 3 the standard deviation is .025. The variation in compost per plate is slightly larger in Cycle 3. For Cutter the standard deviation in Cycle 1 is 0.02. That number falls slightly to 0.16 showing Cutter has slightly less variation in compost per plate for Cycle 3.

A scatter plot of plates and compost shows that there is a general upward trend between the two variables. The more plates used, the more compost is generated. Much of the data is clustered around 340 plates and between 40 and 60 pounds of compost. During Cycle 1, the lines of best fit for both houses are very close. With fewer plates, King produced more compost. The lines intersect at approximately 325 plates and 45 pounds of compost, where Cutter starts creating more compost per plate. In Cycle 3, however, the line of best fit for King lies below the line for Cutter.

The histograms show the distribution of compost per plate for both Cutter and King, but has been divided into Cycle 1 and Cycle 3. The histogram for compost in Cycle 1, shows the highest frequencies around .125-.162 pounds of compost. The data has a small spread and no outliers. King house has the highest frequency of .162 pounds of compost per plate and Cutter has the highest frequency of .130 pounds of compost per plate. Cycle 3 has the highest concentration of rates between 0.13-0.17 pounds of compost per plate. This is a much smaller spread compared to Cycle 1. King is unimodal with a peak around 0.16 and Cutters data has a wider spread without the distinctive peak.

For King and Cutter there is a high correlation between plates and compost for the period of time when tracked. This intuitively makes sense because when there are more plates, it is a signal for more people. When there are more people eating in the dining hall, more waste will be produced.

I ran a regression to look at the effect of the variables poster data and house on the level of compost per plat each night. On average when the poster_date variable is turned on (4/5/20) there is a -8.388e-03 decrease in compost per plate when adjusted for house. The number -2.502e-05 shows the difference between King and Cutter. King house slightly less compost. However, neither of the these small decreases are significant. Additionally, the R^2 value is quite small, 0.02836. The majority of the total change in compost level was no due to the poster date. The regression is set up to only look at the results for Cycle 1 and Cycle 2.