Data Appendix to "Smith Compost Analysis"

Kate Ginder

Contents

1 Appendix description

Your Data Appendix should begin with a brief statement explaining its purpose like the following one.

This Data Appendix documents the data used in "Smith Compost Anlysis". 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 dinning hall.

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

Note: this document structure will require you to re-run steps of your analysis multiple times. If your code takes a long time, please come talk with me about strategies to reduce run time or save earlier results.

2 Raw data

*Each dataset you use will have its own documentation section. The next subsection in this document (Dataset description) is a template. You can copy this section and paste it into your document each time you need to add a section for a new dataset. Note that each line in the Dataset description section **must** end with two spaces.*

This section documents the data sets used in this analysis.

2.1 Dataset description

Citation: Put citation here in APA or other consistent format that you will use throughout the project. Include a hyperlink if applicable.
Date Downloaded: 04/17/2020

Filename: Compost Tracker 3.0.xlsx. **Unit of observation:** amount of compost recorded in dinning 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

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

2.1.3 Variable descriptions

The following data is from two of the Smith College dinning serviouses

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

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

2.1.4 Data import code and summary

Once you've described the variables, enter an R chunk by selecting Code -> Insert Chunk, or Ctrl+Alt+I, give it a name to describe the dataset you are importing. After importing, export a dataframe summary using the command.

```
library(readr)
importable_data <- read_csv("Raw Data Seminar Paper/importable_data.csv")
```

```
## 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()
## )
```

```
View(importable_data)
```

```
Compost_data <- read_csv("Raw Data Seminar Paper/importable_data.csv")  %>%
  rename(king_plates = `#ofplatesKing`,
         king_compost = lbcompostKing,
         cutter_plates = `#ofplatescutter`,
         cutter_compost = lbcompostCutter)  %>%
pivot_longer(contains("_"), names_to = c("house","variable"), names_sep = '_', values_to = "values") %>%
  pivot_wider(names_from = "variable", values_from = "values") %>%
  mutate(date_var = as.Date(dates, "%m/%d/%y")) %>%
  mutate(cycle = case_when(date_var <= as.Date("2020-03-01") ~ 1,
date_var <= as.Date("2020-04-05") ~ 2,
date_var <= as.Date("2020-05-03") ~ 3))
```

```
## 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()
## )
```

```
View(Compost_data)
```

3   Analysis Variables

This section should include a description of all the variables that are used in your final analysis. At the end of the section, you should save all of these variables in the processed_data folder of your repository.

Variables used in the final analysis are

- **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.

```
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   :1.976
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

```r
king_data <- Compost_data %>% filter(house == "king")

summary(filter(king_data, cycle == 1))
```

dates   Dayoftheweek   meal number   house

Length:29 Length:29 Min.   : 1.00 Length:29
Class :character Class :character 1st Qu.: 7.00 Class :character
Mode :character Mode :character Median :14.00 Mode :character
Mean   :14.03
3rd Qu.:21.00
Max.   :28.00
plates compost date_var cycle
Min.   :121 Min.   :17.25 Min.   :2020-02-02 Min.   :1
1st Qu.:210 1st Qu.:32.00 1st Qu.:2020-02-09 1st Qu.:1
Median :333 Median :43.00 Median :2020-02-16 Median :1
Mean   :285 Mean   :41.70 Mean   :2020-02-16 Mean   :1
3rd Qu.:345 3rd Qu.:52.25 3rd Qu.:2020-02-23 3rd Qu.:1
Max.   :370 Max.   :64.25 Max.   :2020-03-01 Max.   :1

```r
summary(filter(king_data, cycle == 3))
```

dates   Dayoftheweek   meal number   house

Length:27 Length:27 Min.   : 2.0 Length:27
Class :character Class :character 1st Qu.: 8.5 Class :character
Mode :character Mode :character Median :15.0 Mode :character
Mean   :15.0
3rd Qu.:21.5
Max.   :28.0
plates compost date_var cycle
Min.   :138.0 Min.   :14.00 Min.   :2020-04-06 Min.   :3
1st Qu.:229.5 1st Qu.:28.75 1st Qu.:2020-04-12 1st Qu.:3
Median :335.0 Median :39.75 Median :2020-04-19 Median :3
Mean   :291.9 Mean   :40.68 Mean   :2020-04-19 Mean   :3
3rd Qu.:352.0 3rd Qu.:51.12 3rd Qu.:2020-04-25 3rd Qu.:3
Max.   :368.0 Max.   :71.25 Max.   :2020-05-02 Max.   :3

```r
cutter_data <- Compost_data %>% filter(house == "cutter")

summary(filter(cutter_data, cycle == 1))
```

```
dates              Dayoftheweek        meal number        house
```

Length:29 Length:29 Min. : 1.00 Length:29
Class :character Class :character 1st Qu.: 7.00 Class :character
Mode :character Mode :character Median :14.00 Mode :character
Mean :14.03
3rd Qu.:21.00
Max. :28.00
plates compost date_var cycle
Min. :133 Min. :17.25 Min. :2020-02-02 Min. :1
1st Qu.:208 1st Qu.:29.25 1st Qu.:2020-02-09 1st Qu.:1
Median :327 Median :46.25 Median :2020-02-16 Median :1
Mean :289 Mean :41.74 Mean :2020-02-16 Mean :1
3rd Qu.:353 3rd Qu.:53.50 3rd Qu.:2020-02-23 3rd Qu.:1
Max. :362 Max. :61.00 Max. :2020-03-01 Max. :1

```r
summary(filter(cutter_data, cycle == 3))
```

```
dates              Dayoftheweek        meal number        house
```
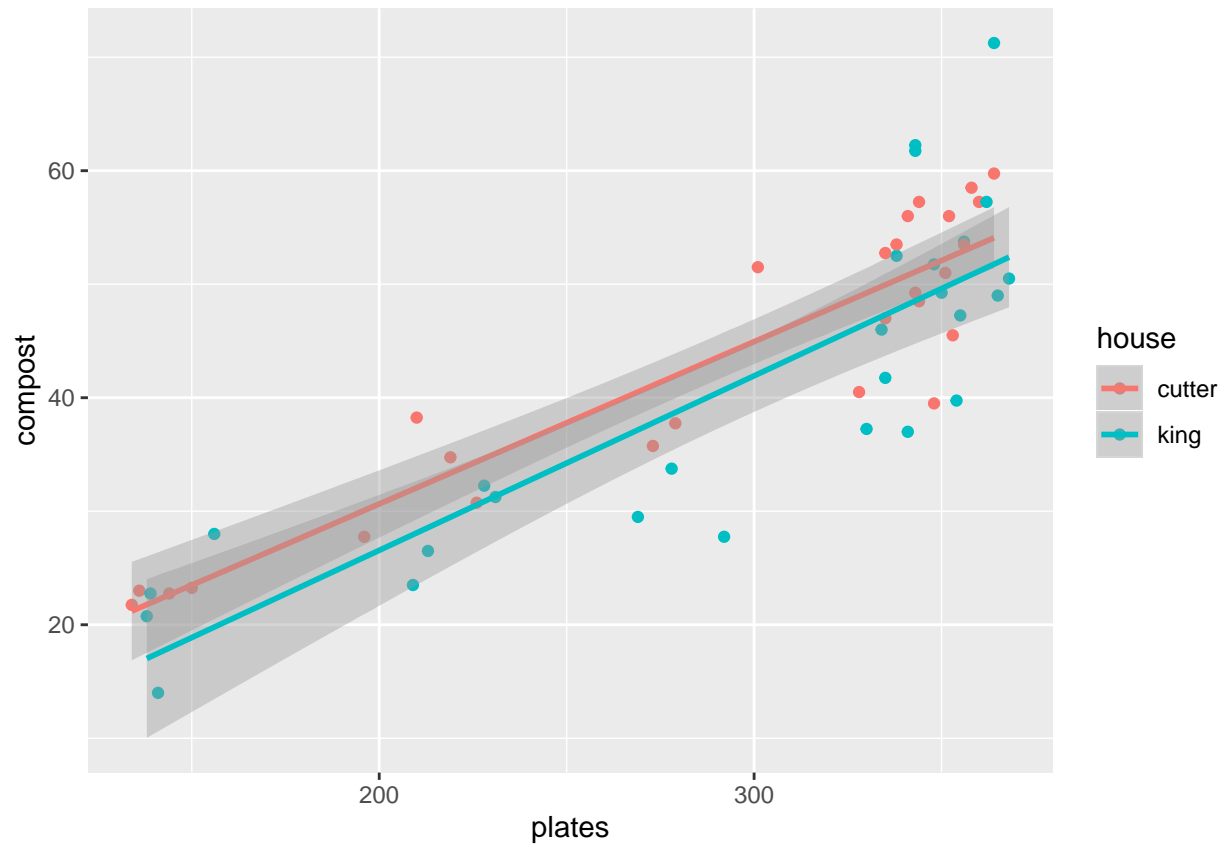
Length:27 Length:27 Min. : 2.0 Length:27
Class :character Class :character 1st Qu.: 8.5 Class :character
Mode :character Mode :character Median :15.0 Mode :character
Mean :15.0
3rd Qu.:21.5
Max. :28.0
plates compost date_var cycle
Min. :134.0 Min. :21.75 Min. :2020-04-06 Min. :3
1st Qu.:222.5 1st Qu.:35.25 1st Qu.:2020-04-12 1st Qu.:3
Median :335.0 Median :47.00 Median :2020-04-19 Median :3
Mean :289.6 Mean :43.44 Mean :2020-04-19 Mean :3
3rd Qu.:349.5 3rd Qu.:53.50 3rd Qu.:2020-04-25 3rd Qu.:3
Max. :364.0 Max. :59.75 Max. :2020-05-02 Max. :3

```r
ggplot(data = filter(Compost_data, cycle == 1), aes(plates, compost, color = house)) + geom_point() + ge
```
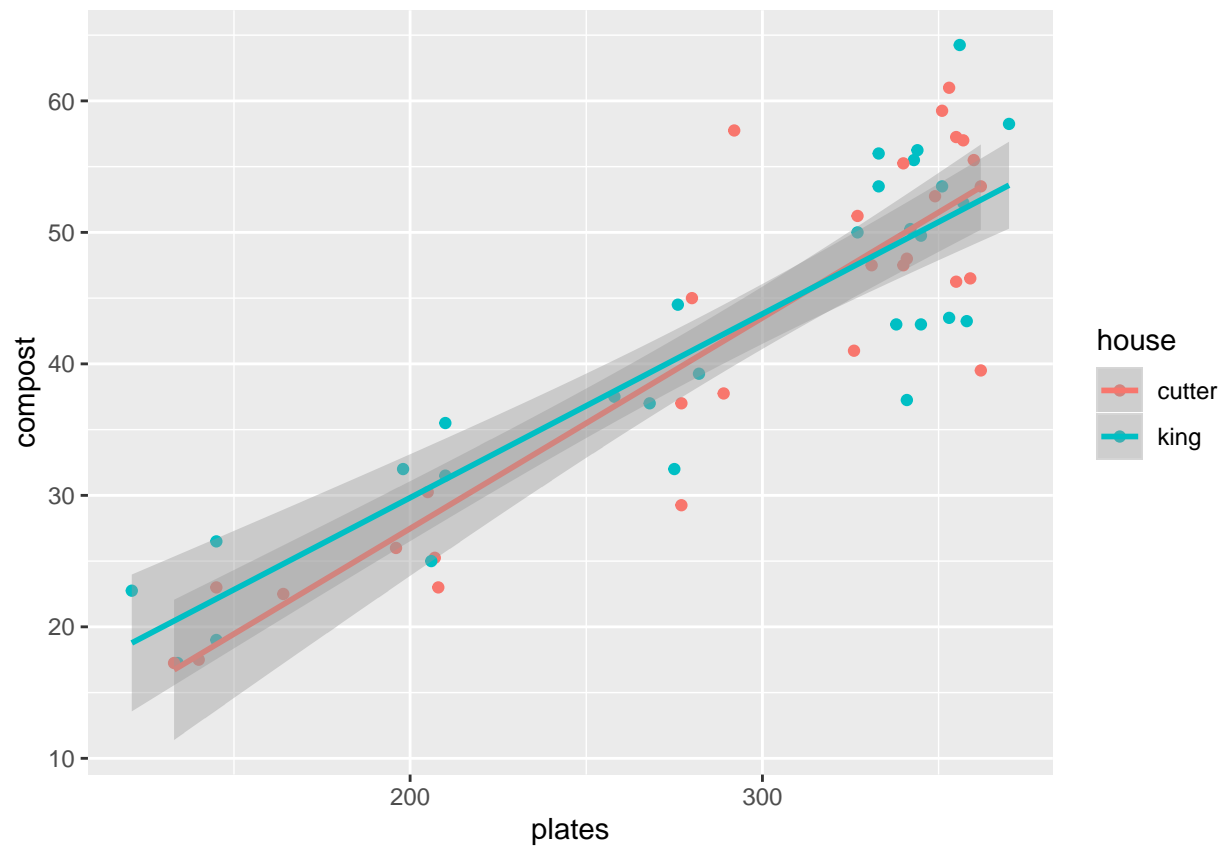
```
ggplot(data = filter(Compost_data, cycle == 3), aes(plates, compost, color = house)) + geom_point() + ge
```

4 Discussion of Data