

Untitled

2024-12-16

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
library(dplyr)
```

```
##  
## Attaching package: 'dplyr'  
  
## The following objects are masked from 'package:stats':  
##  
##   filter, lag  
  
## The following objects are masked from 'package:base':  
##  
##   intersect, setdiff, setequal, union
```

```
library(tidyr)  
library(stringr)  
strawberry <- read.csv("strawberries25_v3.csv") %>%  
filter(Geo.Level %in% c("NATIONAL", "STATE"))  
drop_single_value_cols <- function(df) {  
  single_value_cols <- sapply(df, function(col) length(unique(col)) == 1)  
  df <- df[, !single_value_cols]  
  return(df)  
}  
strawberry <- drop_single_value_cols(strawberry)  
straw_cen <- strawberry %>% filter(Program == "CENSUS")  
straw_sur <- strawberry %>% filter(Program == "SURVEY")  
straw_cen_cleaned <- straw_cen %>%  
separate(Data.Item, into = c("Commodity_Type", "Operation_Measure"), sep = " - ", extra = "merge", fill = "right")  
separate(Commodity_Type, into = c("Commodity", "Type"), sep = ", ", extra = "merge", fill = "right") %>%  
mutate(  
  Type = coalesce(Type, "OTHER"),  
  Operation_Measure = str_trim(Operation_Measure)  
)  
straw_sur_cleaned <- straw_sur %>%  
separate(Data.Item, into = c("Commodity_Market", "Details"), sep = " - ", extra = "merge", fill = "right")  
separate(Commodity_Market, into = c("Commodity", "Market_Type"), sep = ", ", extra = "merge", fill = "right")  
separate(Details, into = c("Measure_Operation", "Unit_of_Measure"), sep = ", MEASURED IN ", extra = "merge", fill = "right")  
mutate(  
  Market_Type = coalesce(Market_Type, "OTHER"),  
  Measure_Operation = str_trim(Measure_Operation),  
  Unit_of_Measure = str_trim(Unit_of_Measure)
```

```

)
straw_sur_cleaned <- straw_sur_cleaned %>%
separate(Domain.Category, into = c("Chemical_Use", "Chemical_Details"), sep = ": ", extra = "merge", fill = "right")
mutate(
  Chemical_Use = str_remove(Chemical_Use, "CHEMICAL, "),
  Chemical_Details = ifelse(Chemical_Use == "NOT SPECIFIED", "NOT SPECIFIED", Chemical_Details)
) %>%
separate(Chemical_Details, into = c("Chemical_Name", "Chemical_Code"), sep = " = ", extra = "merge", fill = "right")
mutate(
  Chemical_Name = str_trim(Chemical_Name),
  Chemical_Code = as.numeric(str_trim(Chemical_Code))
)

```

```

## Warning: There was 1 warning in `mutate()`.
## i In argument: `Chemical_Code = as.numeric(str_trim(Chemical_Code))`.
## Caused by warning:
## ! NAs introduced by coercion

```

```

straw_cen_cleaned <- straw_cen_cleaned %>%
separate(Domain.Category, into = c("Category_Type", "Details"), sep = ": ", extra = "merge", fill = "right")
mutate(
  Details = ifelse(Category_Type == "NOT SPECIFIED", "NOT SPECIFIED", str_trim(Details))
)
straw_sur_cleaned <- straw_sur_cleaned %>%
mutate(State.ANSI = replace_na(State.ANSI, -1))
straw_cen_cleaned <- straw_cen_cleaned %>%
mutate(State.ANSI = replace_na(State.ANSI, -1))
straw_cen_cleaned <- straw_cen_cleaned %>%
mutate(Value = as.numeric(Value)) %>%
drop_na(Value, Year, Category_Type, Details, State.ANSI)

```

```

## Warning: There was 1 warning in `mutate()`.
## i In argument: `Value = as.numeric(Value)`.
## Caused by warning:
## ! NAs introduced by coercion

```

```

straw_sur_cleaned <- straw_sur_cleaned %>%
mutate(Value = as.numeric(Value)) %>%
drop_na(Value, Year, Market_Type, Unit_of_Measure, State.ANSI)

```

```

## Warning: There was 1 warning in `mutate()`.
## i In argument: `Value = as.numeric(Value)`.
## Caused by warning:
## ! NAs introduced by coercion

```

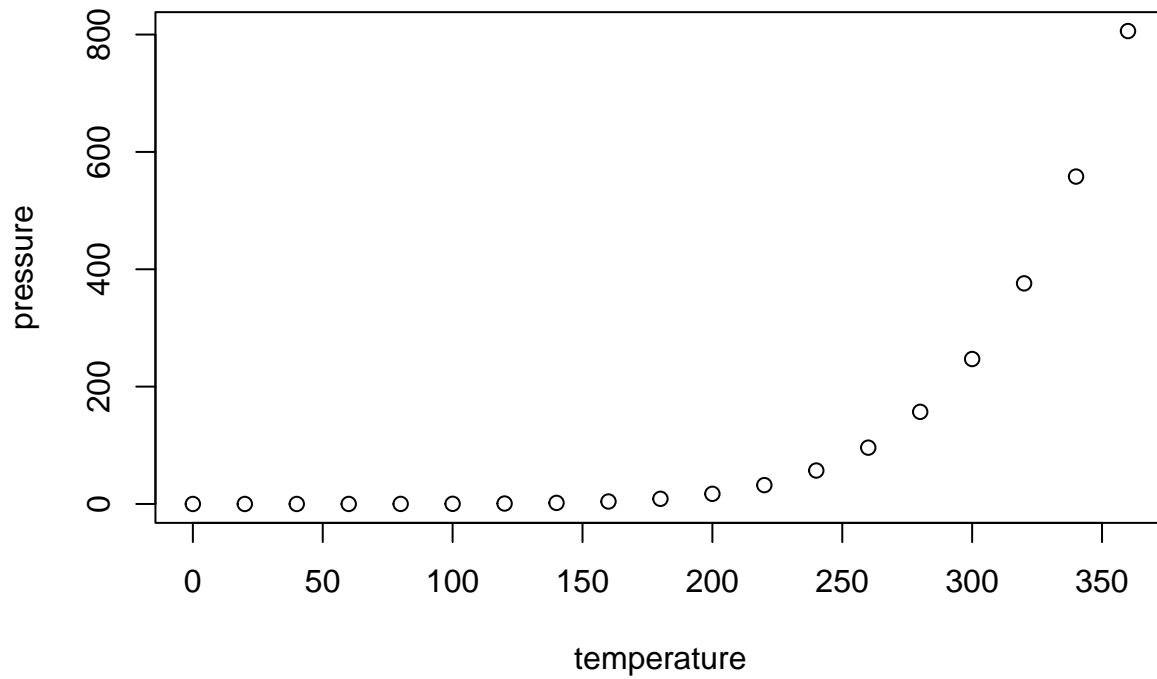
```

model_cen <- lm(Value ~ Year + Category_Type + Details + State.ANSI, data = straw_cen_cleaned)
predicted_cen <- predict(model_cen, newdata = straw_cen_cleaned[is.na(straw_cen_cleaned$Value), ])
straw_cen_cleaned$Value[is.na(straw_cen_cleaned$Value)] <- predicted_cen
model_sur <- lm(Value ~ Year + Market_Type + Unit_of_Measure + State.ANSI, data = straw_sur_cleaned)
predicted_sur <- predict(model_sur, newdata = straw_sur_cleaned[is.na(straw_sur_cleaned$Value), ])
straw_sur_cleaned$Value[is.na(straw_sur_cleaned$Value)] <- predicted_sur
write.csv(straw_sur_cleaned, "straw_sur_cleaned2.csv", row.names = FALSE)
write.csv(straw_cen_cleaned, "straw_cen_cleaned2.csv", row.names = FALSE)

```

Including Plots

You can also embed plots, for example:



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.