ggplot2 with the cereals dataset

Joe Skufca

February 20, 2017

My intention here is not to provide a tutorial on ggplot2. Those already exist. What I did want to show you is how I (conceptually) might have manipulated the dataset we were using for the cereals homework assignments to make a few plots.

Getting the data into R: I will use readr

```
# load the package and read the cereals.csv file from working directory
library(readr)
df=read_csv('cereals.csv') # let function guess the column types
```

```
library(dplyr)
library(ggplot2)
```

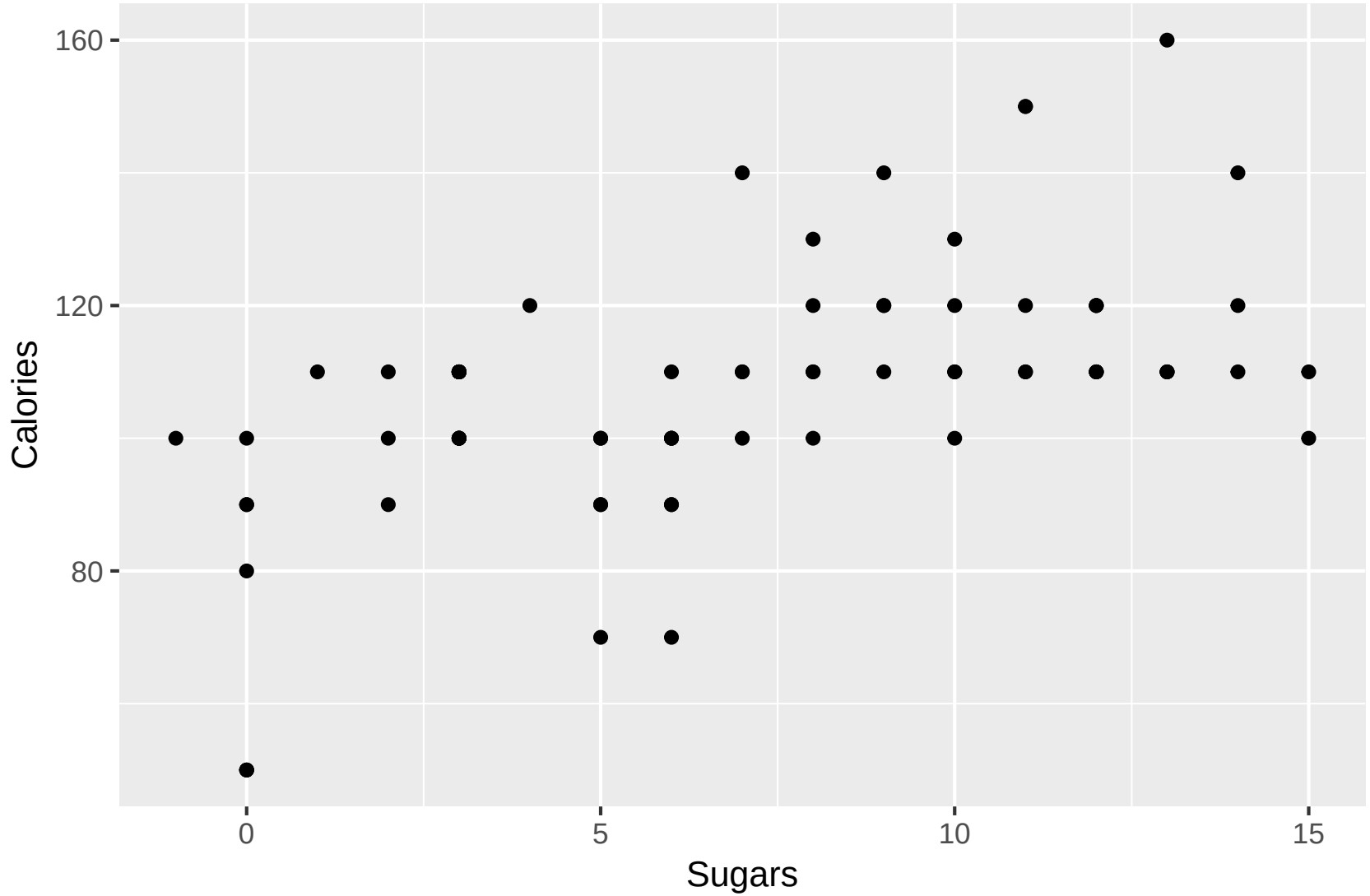
Why ggplot2?

ggplot2 is a bit more complicated for simple plotting, but for more general analysis of a dataset, ggplot2 provides a robust set of static visualizations. *Additionally*, it is built on a conceptual framework, **The Grammar of Graphics**, that matches closely with the theoretical underpinnings of data visualization that we are considering in this course.

A first graph

A simple scatterplot using two of the variables in our dataset:

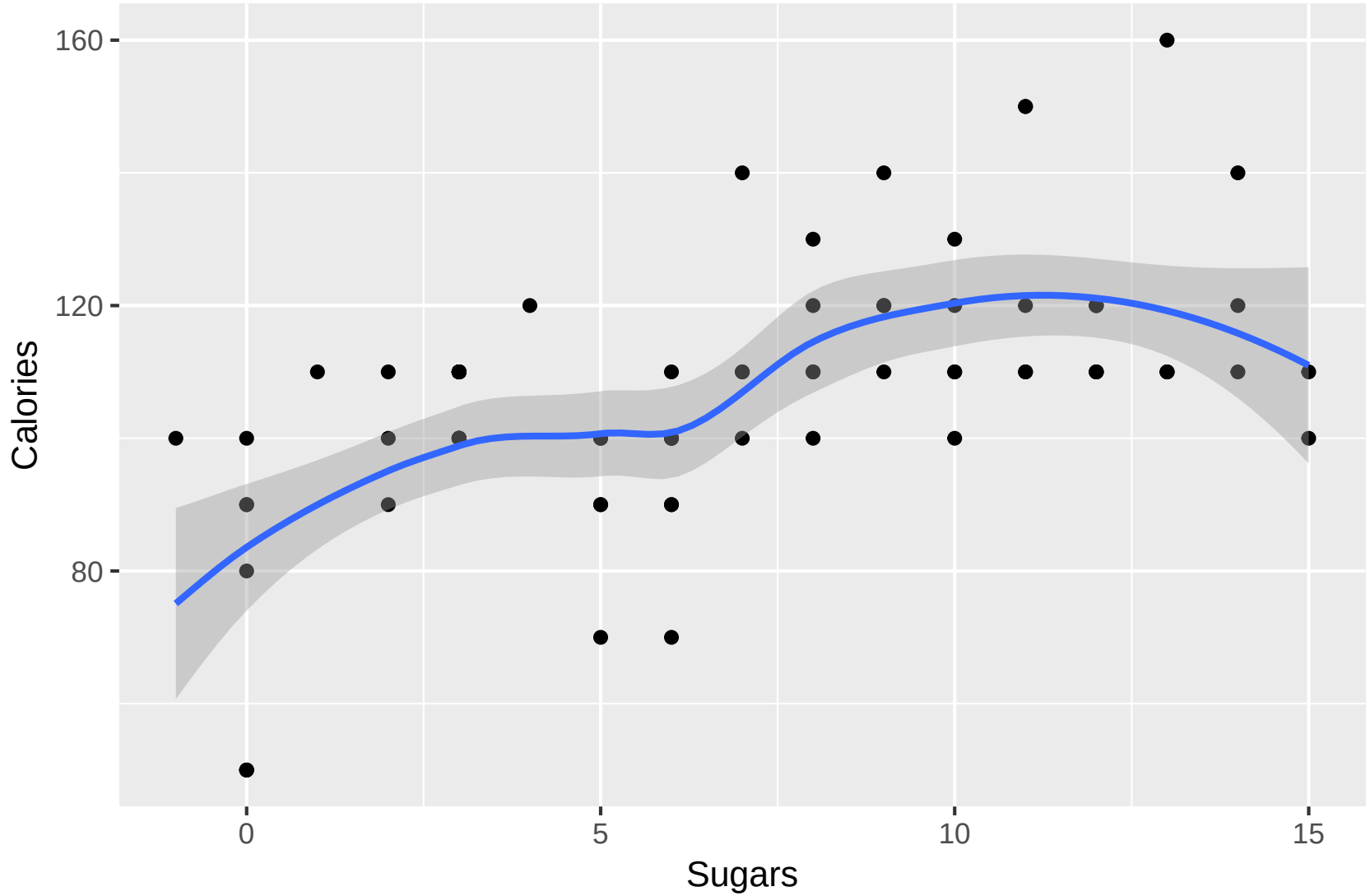
```
p=ggplot(df,aes(x=Sugars,y=Calories))+
  geom_point()
p
```



Manipulation of that plot by adding layers

Adding a smooth curve (based on a statistical model)

```
p=p+geom_smooth()
p
```



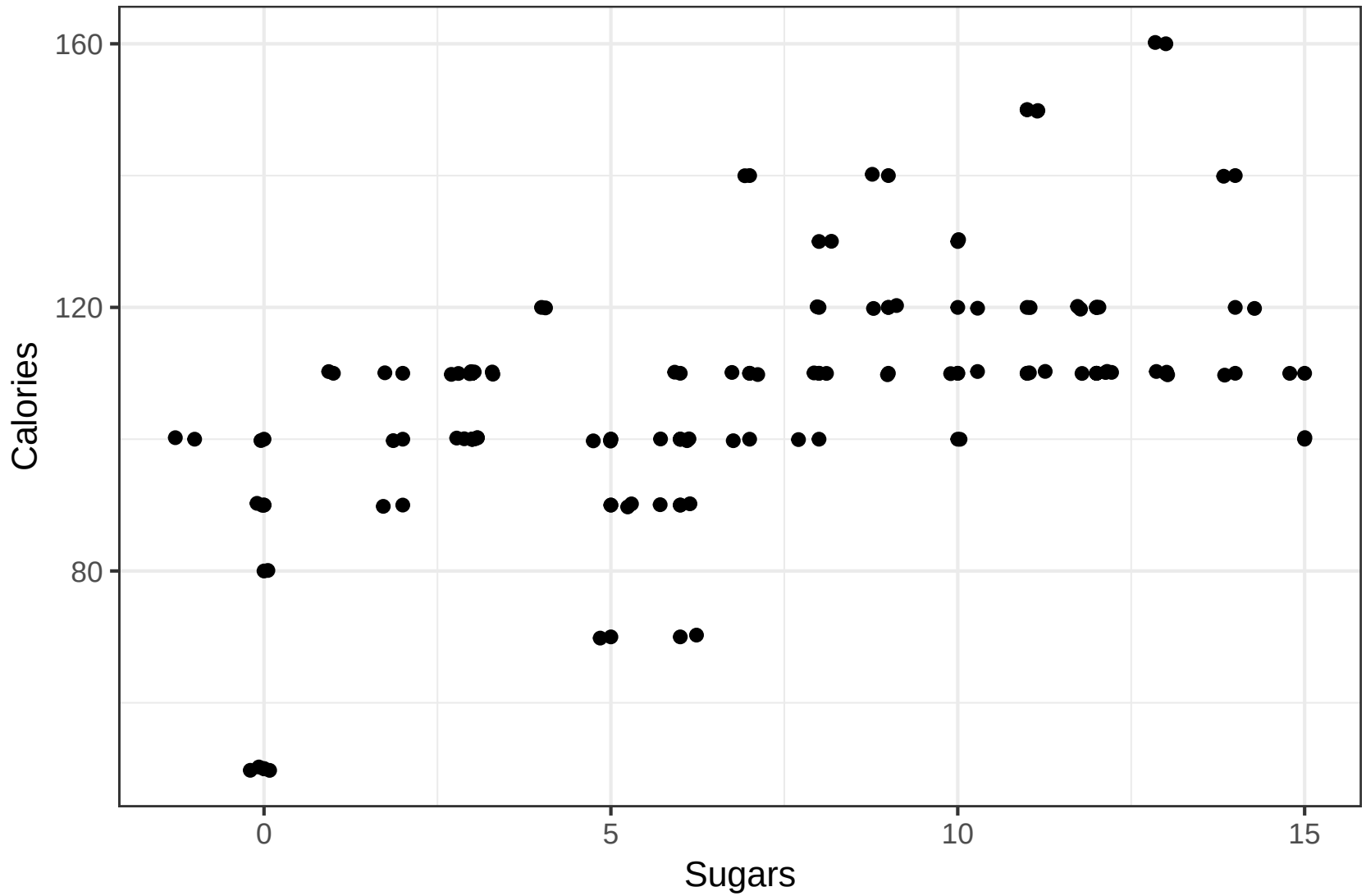
#

Manipulation of that plot by adding a theme Adding a theme manipulates broad characteristics

Multiple layers can be added all at once

Here I also added a jitter effect

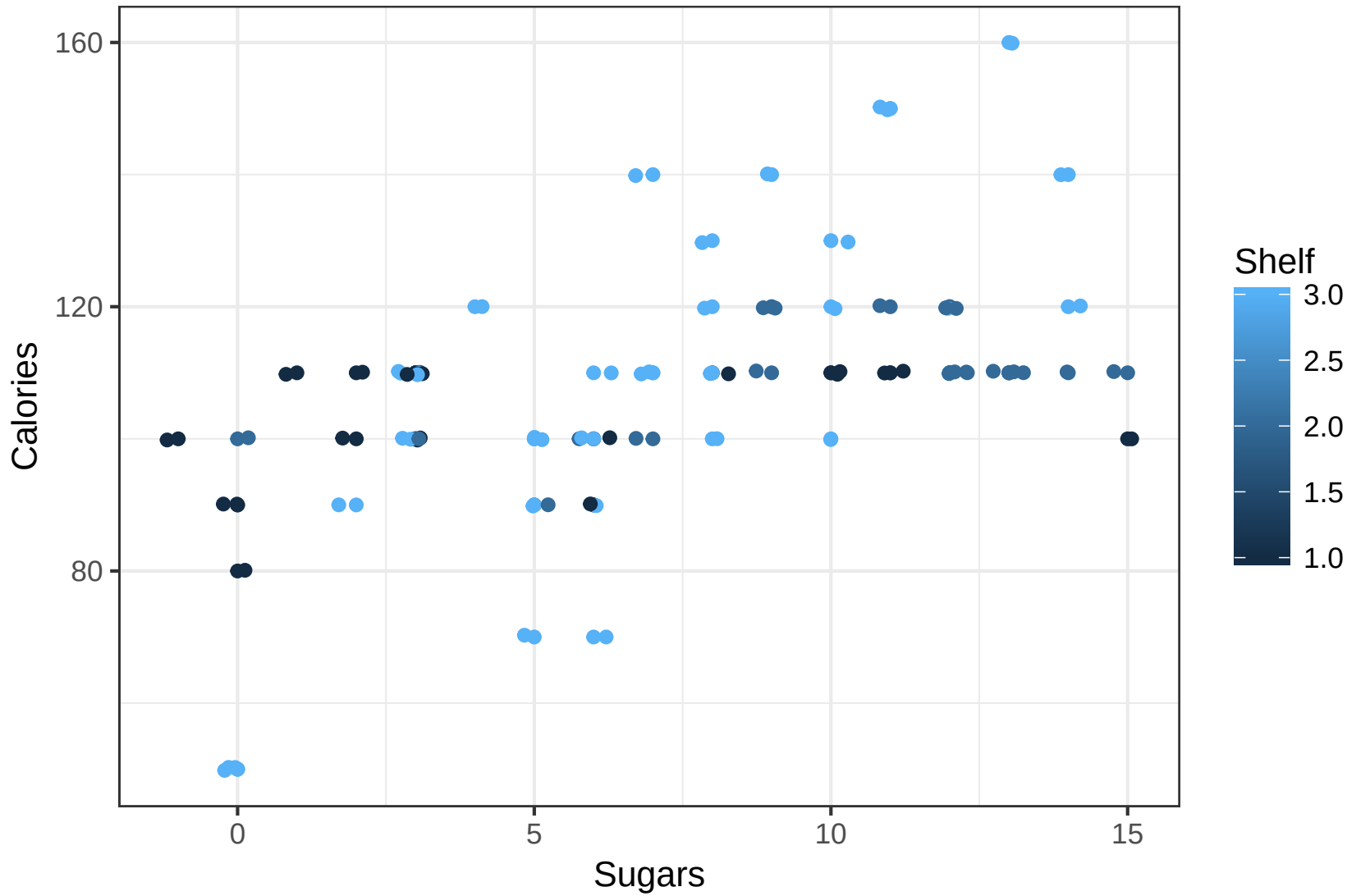
```
p=ggplot(df,aes(x=Sugars,y=Calories))+
  geom_point()+theme_bw()+geom_jitter(width=.3,height=.3)
p
```



aes - aesthetics - assigns a variable to a particular channel

Here I also added a jitter effect

```
p=ggplot(df,aes(x=Sugars,y=Calories,color=Shelf))+
  geom_point()+theme_bw()+geom_jitter(width=.3,height=.3)
p
```

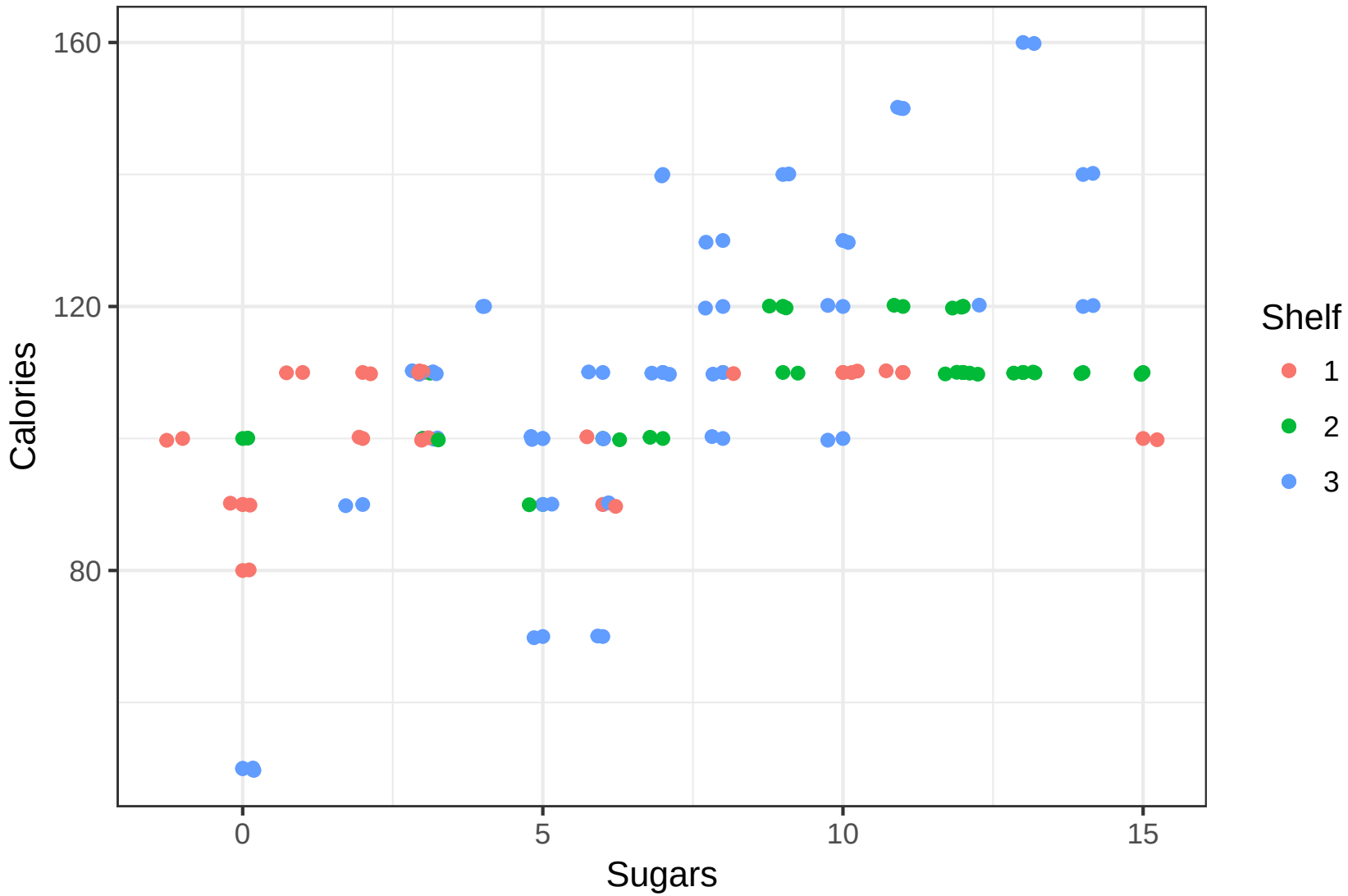


Note

aes - aesthetics - assigns a variable to a particular channel

Lets treat Shelf as categorical

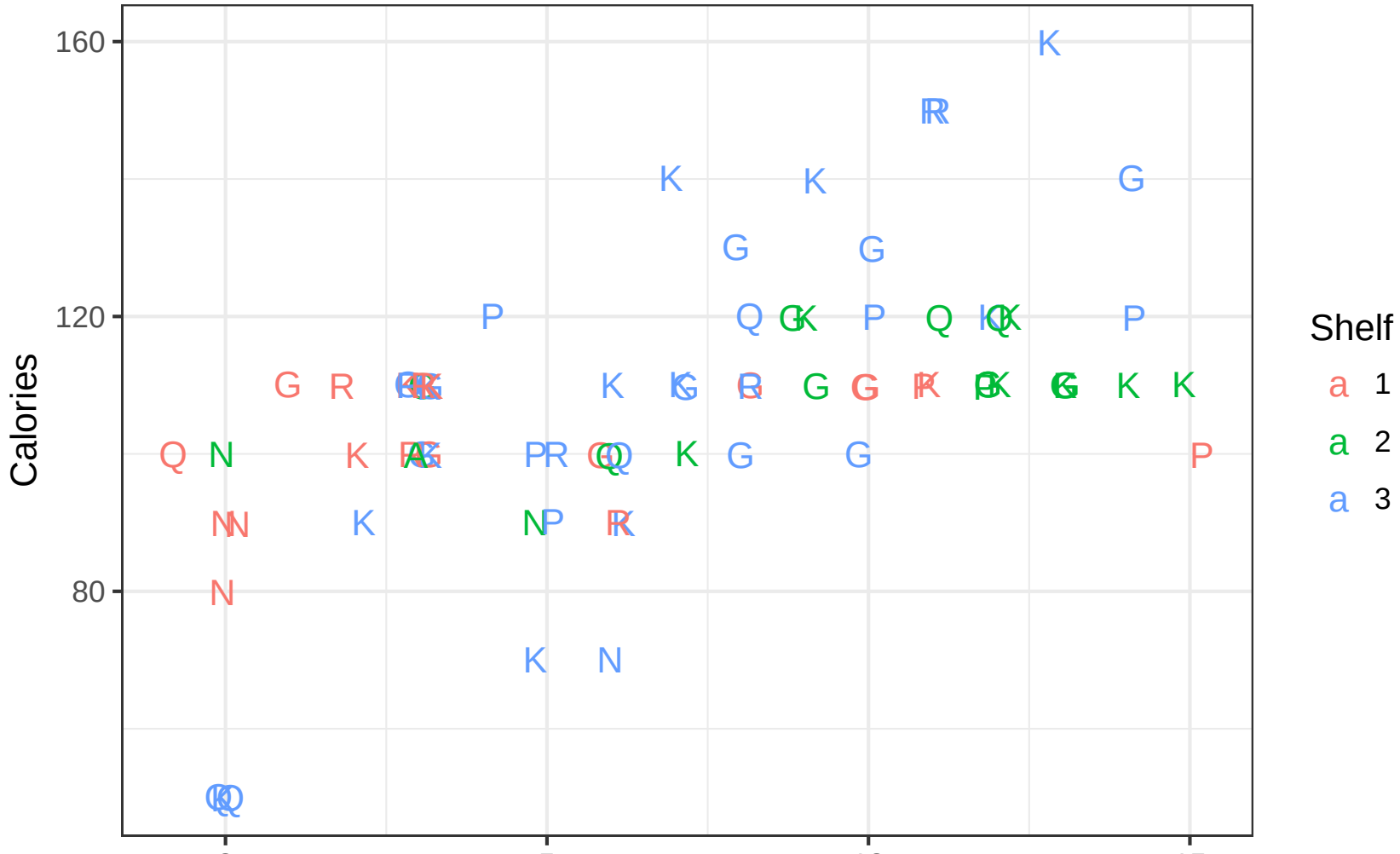
```
p=df %>% mutate(Shelf=factor(Shelf)) %>% ggplot(aes(x=Sugars,y=Calories,color=Shelf))+
  geom_point()+theme_bw()+geom_jitter(width=.3,height=.3)
p
```



geom - geometry - selects a mark, but (sometimes) may also identify an idiom

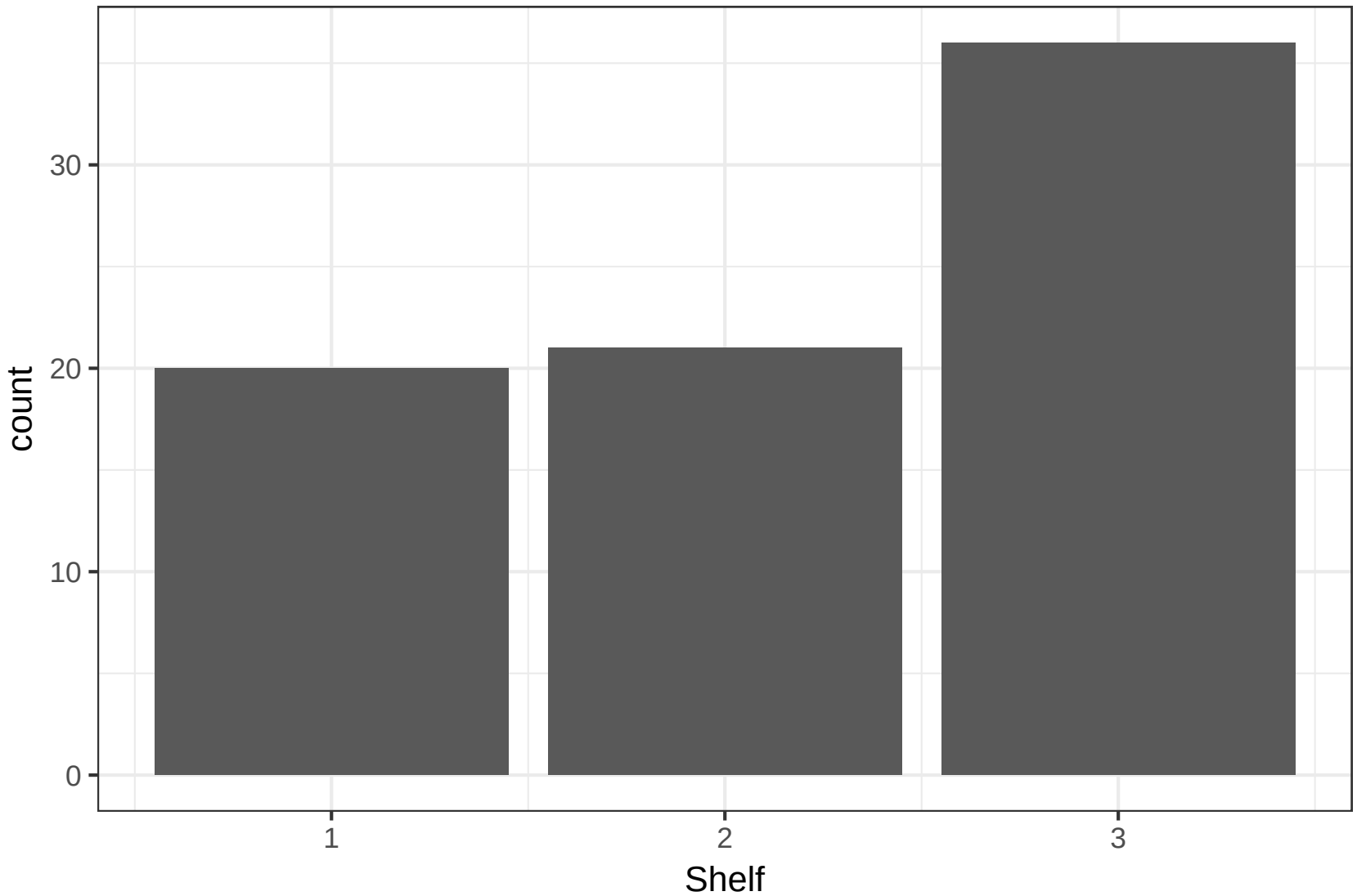
Manufacturer

```
p=df %>% mutate(Shelf=factor(Shelf)) %>% ggplot(aes(x=Sugars,y=Calories,color=Shelf))+
  theme_bw()+
  geom_text(aes(label = Manufacturer), position=position_jitter(width=.2,height=.2))
p
```



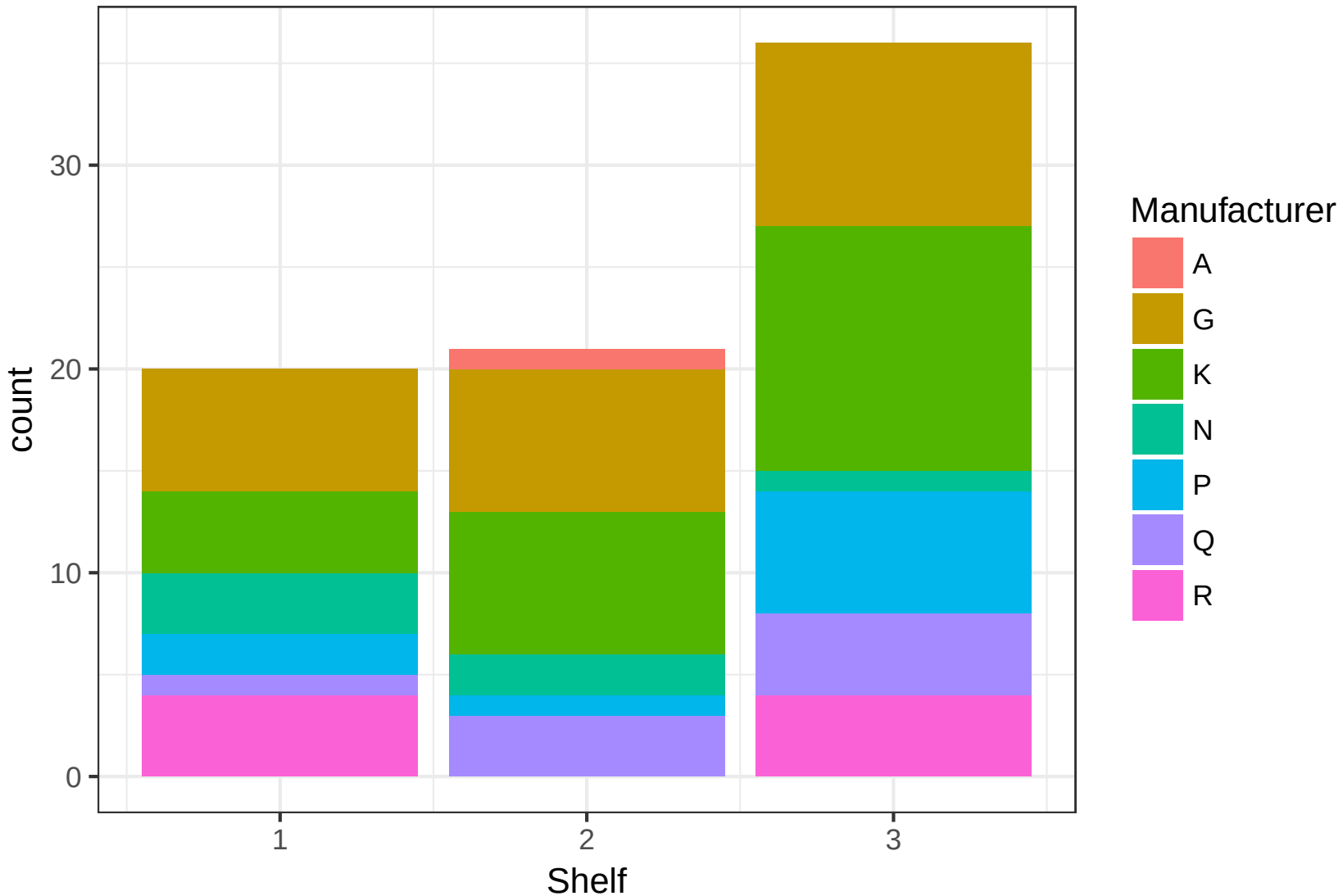
Barcharts - a different geometry

```
p=df %>% ggplot(aes(x=Shelf))+
  theme_bw()+
  geom_bar()
p
```



Stacked Barcharts - “fill” is a channel

```
p=df %>% ggplot(aes(x=Shelf,fill=Manufacturer))+
  theme_bw()+
  geom_bar()
p
```



The scatterplot matrix - because I asked you to

Package GGally adds this functionality

```
library(GGally)
df %>% select(Calories,Protein,Fiber,Carbohydrates) %>% ggpairs()
```