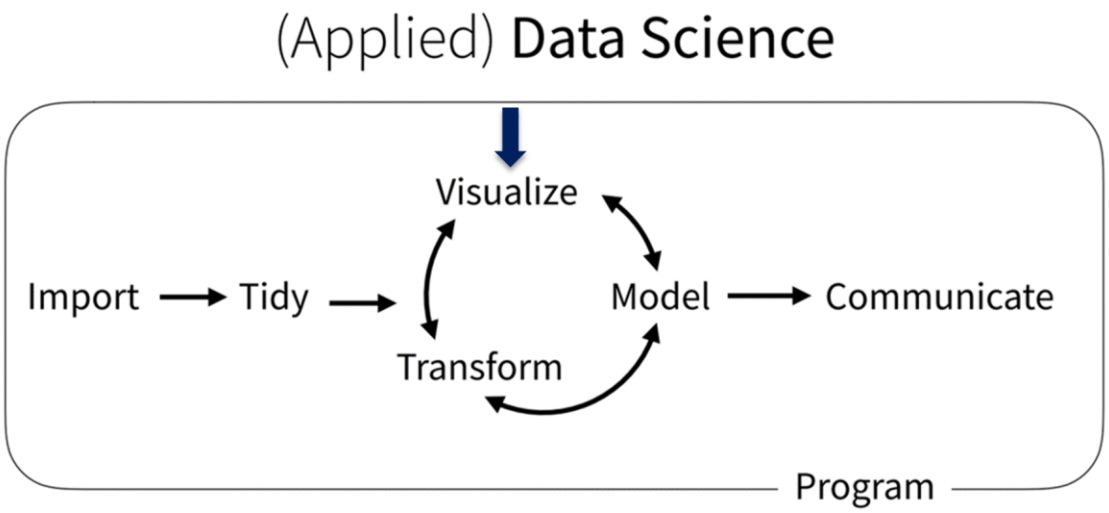
Visualize

Jeff Wesner

September 20, 2019

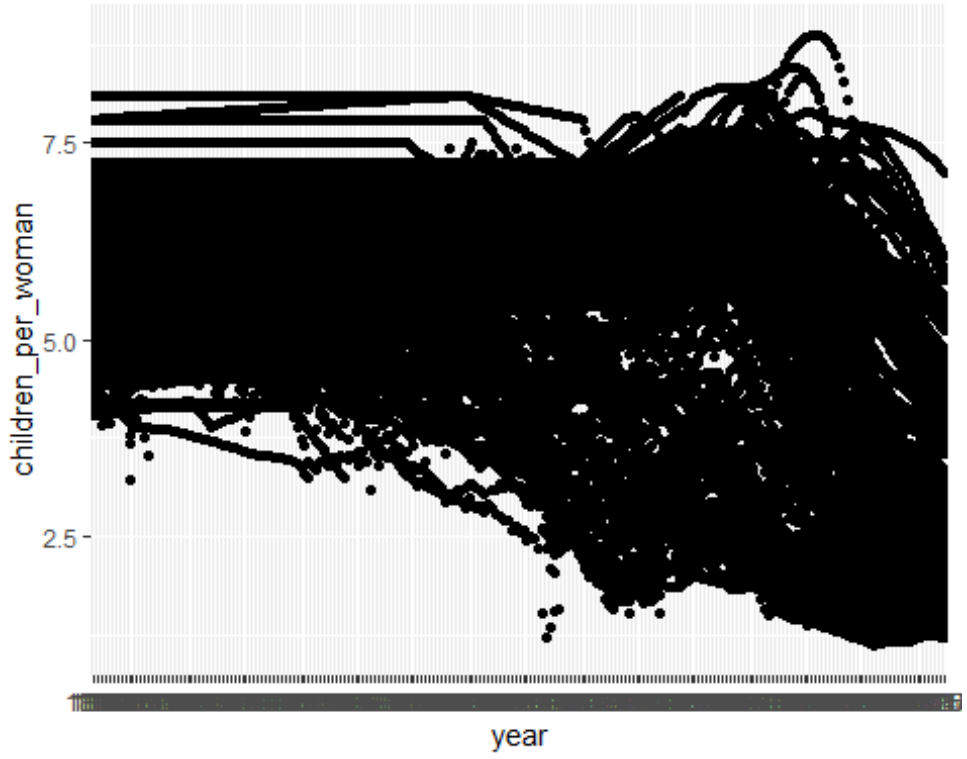


5) Visualize the data

Once your data are in the right format, then it is helpful to visualize them with graphs.

To do that, we'll use *ggplot()*, which is included in the *tidyverse* package. Let's try that by plotting children_per_woman on the y-axis and year on the x-axis.

```
ggplot(total_fertility_long, aes(x=year, y=children_per_woman))+
  geom_point()
```




```
#this code says "plot the dataset total_fertility_long with the following aesthetics: x-axis is year, y-axis is children_per_woman". The second line tells the R how to add the data. In this case we're adding it as a bunch of points, one for each data point.
```

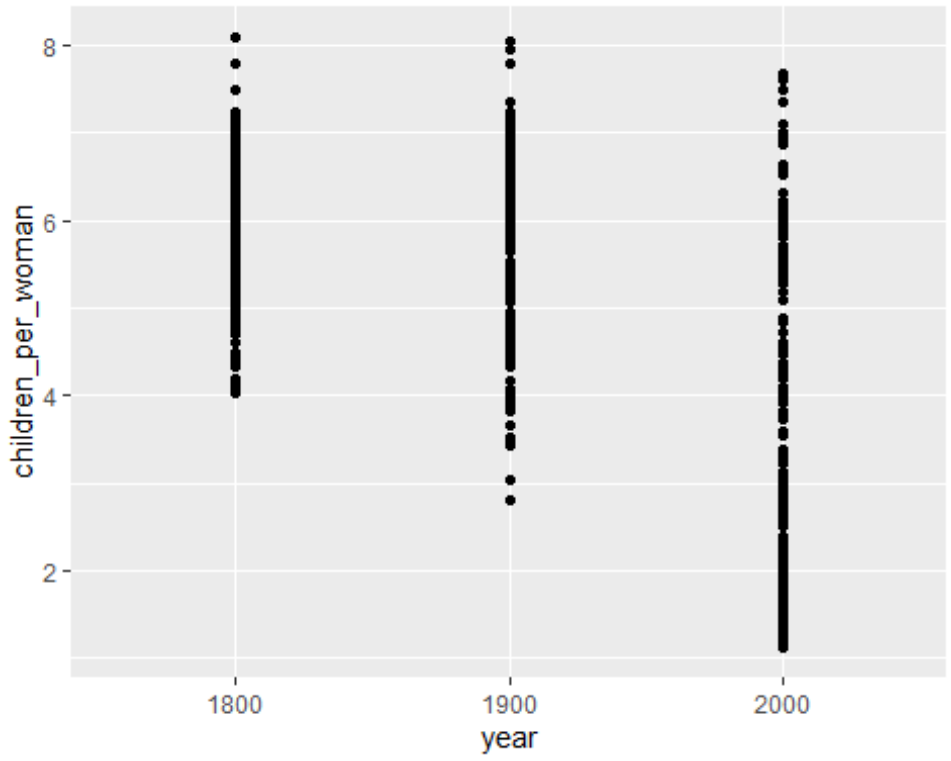
Can you interpret this graph? Sure, it's pretty ugly, but it's a nice first step. Everything after this first step is usually just adding things to make th graph more interpretable. There are a bunch of ways to do this. Here's a cheatsheet of all the things you can do with *ggplot()*. https://www.rstudio.com/wp-content/uploads/2015/03/ggplot2-cheatsheet.pdf

Let's clean it up by only plotting the years 1800, 1900, and 2000. First make a dataset that only contains those years.

```
tot_fert_3years <- total_fertility_long %>%
  filter(year == 1800|year == 1900| year == 2000)
```

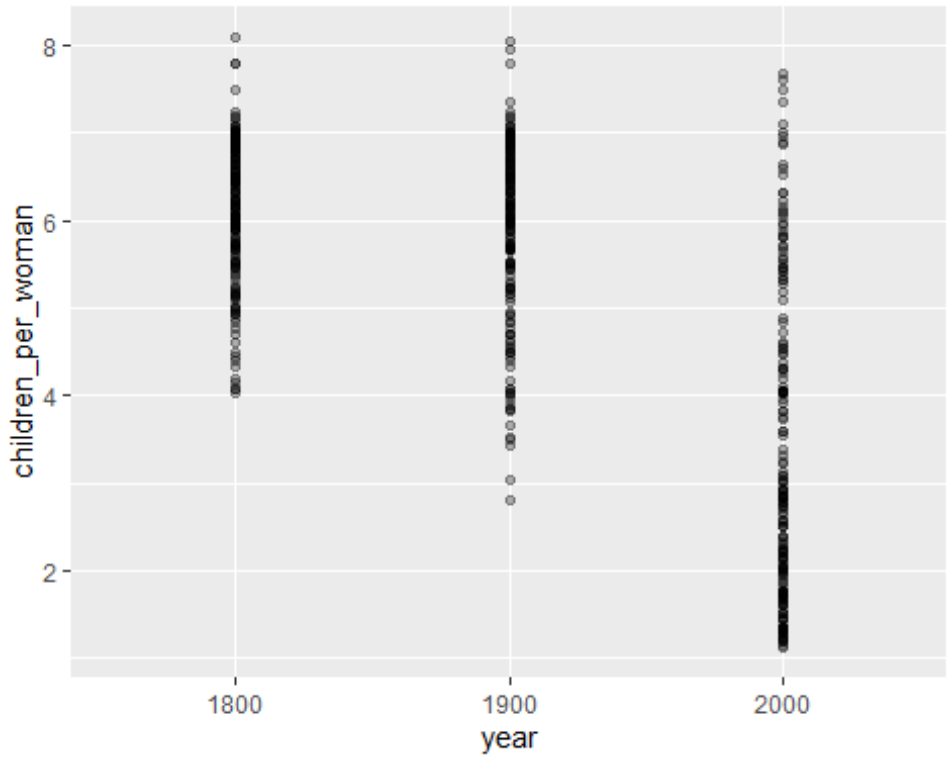
Then use the same plotting code, but now with the tot_fert_3years data.

```
ggplot(tot_fert_3years, aes(x=year, y=children_per_woman))+
  geom_point()
```



That's better, but still a bit hard to see. Let's change the formatting of the dots to make things clearer.

```
ggplot(tot_fert_3years, aes(x=year, y=children_per_woman))+
  geom_point(alpha = 0.3)
```


children_per_woman
8
6
4
2
1800
1900
2000
year