DCF Week 3 Activity

Data and Computing Fundamentals

Popular Names

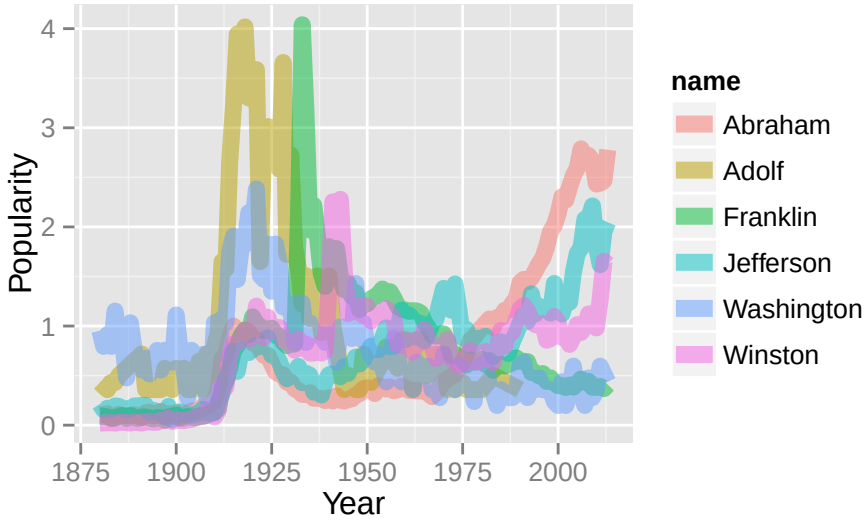
We're going to do some individual and group activities about the *analysis* of data transfiguration tasks as well as their *design*. After constructing a design, and only after, you're going to *implement* the design in R.

In any work, it's important to have a goal and a plan for data table transfigurations. This need not be elaborate.

Fashions in Naming

Objective: create a graph like the following.

```
myNames <- c("Abraham","Franklin","Jefferson","Washington", "Winston", "Adolf")
Results <-
  BabyNames %>%
  filter( name %in% myNames ) %>%
  group_by( name, year ) %>%
  summarise( total=sum(count) ) %>%
  mutate( total=100*total/sum(total))
p <-
  Results %>% ggplot(aes(x=year,y=total,group=name)) +
  geom_line( size=3, alpha=.5, aes(color=name)) +
  ylab("Popularity") + xlab("Year")
plot(p)
```



The raw material you have is the `BabyNames` data set in the DCF package. Write a few example lines in the format of the `BabyNames`, identifying the variable names and typical levels or values.

Point out which variables are categorical. These can potentially be used for defining groups of cases.

First, Individually ...

Step 1. Analyze the graphic to figure out what a glyph-ready data table should look like. Mostly, this involves figuring out what variables are represented in the graph. Write down a small example of the glyph-ready data frame. and how they relate to the variable in the raw materials.

- What variable(s) from the raw data table do not appear at all in the graph?
- What variable(s) in the graph are similar to corresponding variables in the raw data table, but might have been transformed in some way.

Step 2 Consider how the cases differ between the raw input and the glyph-ready table. * Have cases been **filtered** out? * Have cases been grouped and **aggregated/summarized** within groups in any way? * Have any new variables been introduced?

Step 3 Using English, write down a sequence of steps that will accomplish the transfiguration from the raw data table to your hypothesized glyph-ready data table.

Step 4: Confer with your colleagues As a group, compare your different analyses in Steps 1 through 3. Your goal is to develop a consensus for the design in Step 3.

But **FIRST**, each person in turn should present his or her analysis with others in the group **listening**. Don't be shy about saying things like, "But I'm not sure about this," or, "Hearing what person X just said, I realize that this is wrong. I'll explain how it's wrong."

AFTER everyone has presented, start a discussion to develop a consensus about the sequence of operations and the parameters of each (e.g.: What's being filtered out.) Use the whiteboard to keep a record of your consensus.

Step 5: First individually, then as a group. Translate your design, step by step, into R.

Step 6: Implementation Now you can start writing the commands themselves. Do so, try to identify and solve any problems that arise, and make your glyph-ready data.

For graphing, you can use this template:

```
Results %>% ggplot(aes(x=year,y=total,group=name)) +
   geom_line( size=3, alpha=.5, aes(color=name)) +
   ylab("Popularity") + xlab("Year")
```