

# DCF Week 3 Activity

## *Data and Computing Fundamentals*

### Popular Names

The relative popularity of different names for babies varies over the years and decades. In this exercise, you're going to visualize how the popularity of names varies in time.

### Tasks

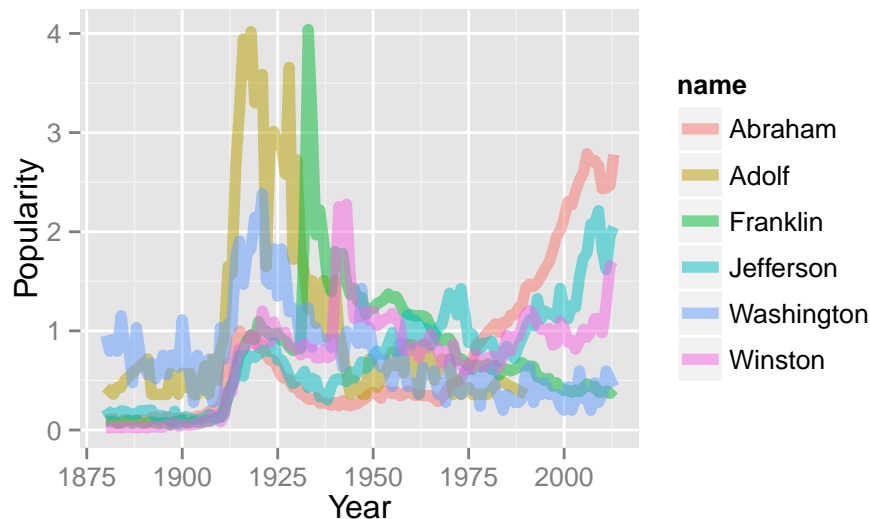
You're going to be working individually first, and then in groups.

- First, undertake an *analysis* of data transfiguration tasks. What kind of data table is the final goal, what is available as the input data. Find a path from input to the final goal in terms of data verbs.
- Second, *design* the steps of the transfiguration process. Describe these in English (if only because your English is a lot better than your R at this stage in your education).
- Third, *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.

### Objective

Create a graph like the following using name of interest to you.



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 a glyph-ready data frame that you think could be used to make something in the form of the graphic.

- 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.

- **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")
```

**Finally ...** A list of Bible-related names is available this way:

```
BibleNames <- fetchData( "DCF/BibleNames.csv" )
```

```
## Retrieving from http://www.mosaic-web.org/go/datasets/DCF/BibleNames.csv
```

Using this data table, and working as a group:

- Make a data table showing the most popular biblenames over all the years.
- Make an informative plot showing the trends over the years of Bible-related names as a proportion of all names.