Summer-To-Do

Daniel Kaplan

May 6, 2014

First day

- Figure out about laptop. Maybe Danny's office machine.
- Install git
- Install preview version of RStudio
- Install latest version of R
- Write and make a screenflow video of how to do this. (3 minutes)
  - Review the whole process with slides, including links
  - Illustrate each step in the process
  - Maybe process this as a tutorial with embedded videos. See HTML5 video tag

Set up a stub Jekyll site for a course

Get direct reading of Google Spreadsheets to work

Look at the `knitr` side-by-side display app.

See the `inst/shiny` directory on the knitr GitHub site.

Make a few example documents that people can play with. E.g. show them a bulleted list, ask them to add an item, to create a sublist, to make a new section. To make a new block and do some R calculation in it.

Graphics

simple map-drawing functions

See `SummerTasks/MapProgram.Rmd[html]`

Look for on-line lovely graphics producing programs

See what format data needs to be in, how to upload it, and create examples of using it. I want to open up the graphics from the pure R functions.

Pick out some compelling interactive graphics from here

A large collection of interactive graphs: http://rcharts.io/gallery/

Build Shiny Apps out of the mosaicManip apps

One directory for each app.

Try the shiny-Rmd for some of them.

Do this for just a few, so you'll be in a position to write new, DCF apps.

Create a Pivot-vs-Fold app

Like the one on p.88 of *Data Transformation_ Skills of the Agile Data Wrangler Presentation.pdf* , but interactive.

Country data

Create a table of synonymous country names and codes. `CaseStudies/Countries`

- `countrycode_data` in `countrycode` package
- The map package (used in 2013) has a table

Scrape the Zip-Code data

They are in URLs like http://www.brainyzip.com/state/zip_newyork.html

Sort out what's useful in `CaseStudies/ZipCodes/zip_codes_states.csv` which is from http://notebook.gaslampmedia.com/download-zip-code-latitude-longitude-city-state-county-csv/. Note that leading zeros have been dropped from the zip code. A data cleaning exercise?

Scrape the County data

http://quickfacts.census.gov/qfd/download_data.html has a database

http://censtats.census.gov/usa/usa.shtml

Bring the Cherry Blossom race up to date

See `CaseStudies/CherryBlossomRace/notes.Rmd`

Federal Election Commission case study

Hadley provides software for reading the file. Do this (out of the project — it's too large to include on github) and then extract a manageable subset.

Crime Data

As in Houston

Break-down crime by day-of-week and offense. Quick start: `ggmap::crime`

FBI data

This one uses:

- `group_by()` see summarize()
- `summarize()` counting events by offense, day, hour, month, etc.
- `filter()` pulling out specific crimes
- `mutate()` turning the number in a day into the percent for that day
- `select()` getting rid of extraneous variables

```
f <- ggmap::crime
ff <- select(f, hour, day, offense, month)
gff <- group_by(ff, hour)
hcount <- summarize(gff, n=n())
ggplot(hcount, aes(x=hour,y=n)) + geom_line(color="blue",width=4)
crimeCount <- group_by(ff,offense,hour) %.% tally()
ggplot(crimeCount,aes(x=hour,y=n,fill=offense)) + geom_bar(stat='identity',position=position_stack(widt
# Do all three types: position_fill, position_dodge, position_stack
ggplot(filter(propCrimeCount,offense%in%c('rape','murder')),
       aes(x=hour,y=p,fill=offense)) + geom_bar(stat='identity',position=position_dodge(width=.9))
# Now reduce each to the proportion
propCrimeCount <- group_by(crimeCount,offense) %.% mutate(p=n/sum(n))
ggplot(filter(propCrimeCount,offense%in%c('rape','murder')),
       aes(x=hour,y=p,fill=offense)) + geom_bar(stat='identity',position=position_dodge(width=.9))
```

European Temperature data

For fields and mutation, wide, long, ...

http://www.datasciencecentral.com/group/resources/forum/topics/importing-100-years-of-climate-change-into-r

US Health Data

Data site NEJM article

Apps for data operations

Mutate

- Fixed set of example data -or- Pick a data set and create the mutation you want by writing in the expression.
- Show the head of the mutated data and the new column alongside.
- Show the statement that will do the job.
- Some quiz exercises.
  - Quiz results: Get the userID, then code the table of results for them to send in. Or, can we arrange to send it automatically?

Join

- Selection of kind of join
- Have a list of example data types, and
- Selection of tables to join from the local disk
- Choose kind of join
- Choose vars to join on. Choice from table A -> Choice from table B

Set up a course web site with Jekyll

Resources, syllabi (perhaps multiple) that refer to the resources.