Basic Computing 2 in-class activity — Revisiting Fujita's study of the 1974 tornado superoutbreak*

Peter Carbonetto *University of Chicago*

We will re-examine data from T. T. Fujita's influential study of the 1974 tornado 'superoutbreak.' Our aim will be to apply the practices we learned in the first activity toward a more complex analysis, then efficiently repeat the same complex calculations across multiple data sets.

Instructions

Locate the "basic_computing_2" folder on your computer. Within that folder, you should see an R code file, `tornadoes.R`. Open this file in RStudio. Then set your working directory to the "basic_computing_2" folder. It is also Good Practice to start this exercise with a clean workspace.

Aims

T. T. Fujita (1920–1998) was a professor at the University of Chicago. He is perhaps best known for the development of the Fujita (F) classification of tornado intensities. The F scale divides tornado wind speeds into 6 classes, labeled F0 up to F5. Several important breakthroughs came from his detailed study of the tornado superoutbreak of 1974. Here we will develop R code to analyze data from the Severe Weather Data Inventory (SWDI) and understand the importance of the 1974 tornado superoutbreak in tornado climatology.

Run analysis, and review code. Our starting point is some code for analyzing the SWDI storm event data from 1974. The code is in `tornadoes_1974.R`, and is reproduced here for convenience:

```
library(readr)
library(ggplot2)
storms <- read_csv("StormEvents_details-ftp_v1.0_d1974_c20210803.csv.gz")
class(storms) <- "data.frame"
storms <- transform(storms,date = as.Date(paste(1974,BEGIN_YEARMONTH - 197400,
                                                BEGIN_DAY,sep = "-")))
storms <- transform(storms,dayofyear = as.numeric(format(date,"%j")))
tornadoes <- subset(storms,EVENT_TYPE == "Tornado")
dat <- as.data.frame(table(tornadoes$dayofyear))
names(dat) <- c("day","count")
dat <- transform(dat,day = as.numeric(as.character(day)))
p <- qplot(x = dat$day,y = dat$count,geom = "line",
           xlab = "day",ylab = "events") +
  theme_classic()
print(p)
```

*This document is included as part of the Basic Computing 2—Introduction to R tutorial packet for the BSD qBio Bootcamp, University of Chicago, 2021. **Current version**: September 07, 2021; **Corresponding author**: pcarbo@uchicago.edu.

This particular R code file is a standalone "program" or "script"—that is, the code can be run inside an empty R environment to produce a useful result:

```
source("tornadoes_1974.R")
```

Our first aim is to understand, at least at a high level, the result that was produced, and how the code generated this result. We will add comments to the code to remind us of the logic behind the individual steps of the analysis.

Create a more flexible data analysis script. Was there something special about the weather in 1974? Here we will compare against the tornado patterns in 1975 and 1976. To do this, we will need to adapt the existing code to work with data from other years. Copy the code to a new file, `tornadoes.R`. Then we will revise the code to make it more flexible. To get you started, here's a suggestion for the first few lines of the new script:

```
library(readr)
library(ggplot2)
year <- 1974
filename <- paste0("StormEvents_details-ftp_v1.0_d",year,"_c20210803.csv.gz")
```

Test your script on the 1974, 1975 and 1976 data by setting "year" to 1974, 1975 or 1976. Make adjustments to your code if you get an error or if the results do not look right.

Implement the analysis as a function. Let's now take this idea one step further: suppose we would like to rerun this same analysis on many more years. At some point it would be tedious to reuse the script. Therefore, we will design a *function* to automate the analysis, building on our flexible analysis script (this isn't the only way to automate this analysis, but designing a function has some other important benefits). Let's assume the inputs to your function are the year and filename, and the output is a plot object returned by `qplot` (and choose a memorable name for the function). Save your function in a file called `tornado_functions.R`. Then try using your new function to analyze the three data sets. Also please add a few helpful comments to `tornado_functions.R` to explain what the function does.

Bonus: Download CSV files for other years from the NOAA's SWDI website and try applying your function to these data.

Currently, our data analysis gives only a very broad picture of tornado patterns in a given year—one might be interested, for example, in a finer grained patterns stratified by tornado intensity (these data are in the "TOR_F_SCALE" column). What other aspects might we consider in our analysis? What inputs would you add to your function to automate an analysis that allows for specifying these additional criteria?