Preparing For Analysis

Before jumping into data retrieval and analysis, we need to setup an area where we can organize all our input data, analysis scripts, output (visualizations, reports and/or data) and any supporting documentation. For the purposes of this chapter, we’ll be using the following directory structure:

/book/ch3

|-*R*

|-data

|-docs

|-output

|-python

|-support

|-tmp

Like most elements of programming, there is no one, true way to setup this structure, but you should strive to find one that works for you and stick with it. A great way to do that is to take a lesson from modern web framework builders and use a simple setup script that builds the structure for you:

Sample analysis preparation script

#!/bin/sh

#

# prep: prep analytics directory structure

#

# usage: prep DIRNAME

#

DIR=$1

if [ ! -d "${DIR}" ]; then

mkdir -p ${DIR}/*R* \

${DIR}/data \

${DIR}/docs \

${DIR}/output \

${DIR}/python \

${DIR}/support \

${DIR}/tmp

> ${DIR}/readme.md

ls -lR ${DIR}

else

echo "Directory "${DIR}" already exists"

fi

You now only need to type “prep NAME” whenver you want to start a new project (so, for this project, “prep ch3”). As you develop your own styles and patterns, you can expand this script to include the generation of various templates and even initialization of source code repositories. Once the structure is in place, it’s time to retrieve, explore and analyze some data.