R Workshop Setup Instructions

R and RStudio

The workshop will include a mix of lecture and hands-on demonstrations. In order to follow along, make sure you bring a laptop with R and RStudio installed. R and RStudio are separate downloads and installations. You need R installed before you install RStudio.

Download and install R for Windows (https://cran.r-project.org/bin/windows/base/) or Mac (http://cran.r-project.org/bin/macosx/).

Download and install RStudio Desktop: https://www.rstudio.com/products/rstudio/download/

Note: If you have previously installed R and RStudio, please confirm that you have R >= version 3.4.0 and RStudio >= version 1.0.0. To confirm this, open up RStudio. In the upper left-hand navigation menu, click "About RStudio" to see its version. Then in the console panel, type `R.version` to see which major/minor releases of R you have installed.

Packages

We will need to install several core packages needed for this lesson. Launch RStudio (RStudio, not R itself). Ensure that you have internet access, then copy and paste the following commands, one-at-a-time, into the console panel (usually the lower-left panel, by default) and hit the Enter/Return key:

```
install.packages("dplyr")
install.packages("readr")
install.packages("tidyr")
install.packages("ggplot2")
install.packages("pryr")
install.packages("microbenchmark")
install.packages("data.table")
install.packages("nycflights13")
```

If you receive an error message when trying to install any particular package, please make note of which one(s) you had trouble with.

A few notes:

- Commands are case-sensitive.
- You must be connected to the internet.
- Even if you've installed these packages in the past, do re-install the most recent version. Many of these packages are updated often, and we may use new features in the workshop that aren't available in older versions.
- If you're using Windows you might see errors about not having permission to modify the existing libraries – disregard these. You can avoid this by running RStudio as an administrator (right click the RStudio icon, then click "Run as Administrator").

Check that you've installed everything correctly by closing and reopening RStudio and entering the following command at the console window (don't worry about any messages that look something like `the following objects are masked from ...`, or `Warning message: package ... was built under R version ...`).

```
library(dplyr)
library(readr)
library(tidyr)
library(ggplot2)
library(pryr)
```

```
library(microbenchmark)
library(data.table)
library(nycflights13)
```

This may produce some notes or other output, but as long as you don't get an error message, you're good to go. If you get a message that says something like: `Error in library(somePackageName) : there is no package called 'somePackageName'`, then the required packages did not install correctly. Make sure you note the names of the package(s) that you had trouble installing. Instructors will be available 30 minutes before class to help resolve any outstanding installation issues.

Data

Please make sure you have the necessary data downloaded to follow along with in class examples.

Brauer yeast gene expression data

Download: https://workshops.somrc.virginia.edu/data/brauer2007_tidy.csv

Cleaned up version of a gene expression dataset from Brauer et al. Coordination of Growth Rate, Cell Cycle, Stress Response, and Metabolic Activity in Yeast (2008) *Mol Biol Cell* 19:352-367. This data is from a gene expression microarray, and in this paper the authors are examining the relationship between growth rate and gene expression in yeast cultures limited by one of six different nutrients (glucose, leucine, ammonium, sulfate, phosphate, uracil). If you give yeast a rich media loaded with nutrients except restrict the supply of a *single* nutrient, you can control the growth rate to any rate you choose. By starving yeast of specific nutrients you can find genes that:

1. Raise or lower their expression in response to growth rate. Growth-rate dependent expression patterns can tell us a lot about cell cycle control, and how the cell responds to stress. The authors found that expression of >25% of all yeast genes is linearly correlated with growth rate, independent of the limiting nutrient. They also found that the subset of negatively growth-correlated genes is enriched for peroxisomal functions, and positively correlated genes mainly encode ribosomal functions.
2. Respond differently when different nutrients are being limited. If you see particular genes that respond very differently when a nutrient is sharply restricted, these genes might be involved in the transport or metabolism of that specific nutrient.

Gapminder

Download: https://workshops.somrc.virginia.edu/data/gapminder.csv

Excerpt from the Gapminder data. The dataset contains information on life expectancy, GDP per capita, and population by country. Variables include country, continent, year, life expectancy at birth, total population, and per-capita GDP in units of [international dollars](internationaldollars.

National Health and Nutrition Examination Survey (NHANES)

Download: https://workshops.somrc.virginia.edu/data/nhanes.csv

NHANES is a research program designed to assess the health and nutritional status of adults and children in the United States. The survey is one of the only to combine both survey questions and physical examinations. The interview includes demographic, socioeconomic, dietary, and health-related questions. The physical exam includes medical, dental, and physiological measurements, as well as several standard laboratory tests. NHANES is used to determine the prevalence of major diseases and risk factors for those diseases.

This dataset is small slice of the NHANES data, with a handful of variables from the 2011-2012 survey years on about 5,000 individuals.