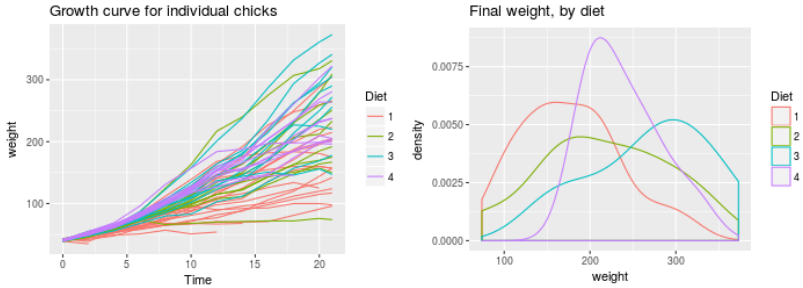
Data Visualization in R

An introduction to the tidyverse and plotting data in R

Workshop Overview

Goals:

- Introduce R as a computing language, installing and loading packages, and markdown
- Learn to load and preview data
- Formatting and cleaning data using `dplyr`
- Create beautiful and informative plots using `ggplot2`



What is R?

Ryerson University

Officially: R is a programming language and free software environment for statistical computing and graphics supported by the R Core Team and the R Foundation for Statistical Computing. It is widely used among statisticians and data miners for developing statistical software and data analysis

Unofficially: a fancy calculator that can make pretty plots

Why do we use it?

- easy to learn and intuitive to write
- fast and powerful with respect to large-ish data sets
- open source so many people contribute to the language
- creates aesthetically pleasing graphs

Getting Started

Ryerson University

1. Installing R
   - Google "R" and go to first link
   - Find your nearest CRAN mirror and download the executable
2. Install the R Studio IDE (makes coding nicer)
   - Google "R studio" and go to first link
   - Products > RStudio > Download RStudio for Desktop
3. Understanding the environment
4. Installing vs. loading packages

Tidyverse

Ryerson University

A collection of R packages designed for data science which share common conventions

So what? These packages are highly optimized and provide a consistent "grammar" making it easy to work with all of them at once

Important packages:

- `readr` - data parser for reading in various files types — Cheatsheet
- `dplyr` - grammar for data manipulation. — Cheatsheet
- `ggplot2` - based on the layered grammar of graphics which provides a dividable framework for plotting data — Cheatsheet