Data Import

- Data files should idealy be set up as text or CSV (comma-separated values) files with rows as cases and columns as variables. Data sets for this course will be found in these formats on Blackboard. Save them to you desktop and use the functions **`read.table()`** or **`read.csv()`** to import them into R as data frames. This will require a complete path to the file's location on your computer.
- Finding the path to your file can be annoying. Here's a convenient alternative: the *Import Dataset* button in the "Environment" window of RStudio. (Look it up!) This starts a wizard that guides you through the process. In the end, the wizard sends a command to your console that contains the path and any options that you specified. Don't forget to copy this command back to your script! This way, next time you run the analysis from your script, the file is imported automatically.
- Try to import a text file. First download the data file "lnyxhare.txt" from Blackboard. This file contains the Canadian lynx and snowshoe hare pelt-trading records of the Hudson Bay Company, starting in 1852. (This data is dowloaded from https://github.com/bblais/Systems-Modeling-Spring-2015-Notebooks/blob/master/data/). How to do this depends on the browser that you're using. Possibly you have to click on the file to open it in the browser, and then right-click to choose *Save As* .... Then either use the data import wizard or use the following code:

```
> lynx_hare <- read.table("C:/.../lynxhare.txt",
     header = TRUE, sep = ";")
> head(lynx_hare) # The function head() returns the first 7 rows of the
    dataset.
```

```
  year  hare lynx
1 1852 80000 2174
2 1853 80000 2106
3 1854 90000 3021
4 1855 69000 4754
5 1856 81000 7324
6 1857 95000 8197
```

```
> str(lynx_hare)
```

```
'data.frame':   88 obs. of  3 variables:
 $ year: int  1852 1853 1854 1855 1856 1857 1858 1859 1860 1861 ...
 $ hare: int  80000 80000 90000 69000 81000 95000 71000 28000 18000 19000 ...

 $ lynx: int  2174 2106 3021 4754 7324 8197 6913 4772 2383 1540 ...
```

- Import a CSV file (you can find the data file "lynxhare.csv" on Blackboard):

```
> lynx_hare <- read.csv("lynxhare.csv",
   sep = ";")
```

- If you need to import data from an Excel file, you can simply save the Excel sheet as a CSV file. In Excel, use *Save As* and select CSV or comma-separated

values. Then use the steps outlined above to import the CSV file. Now let us explore our data a little bit. For example, what are the mean and standard deviation of hare population? We can easily find out about this by:

```
>mean(lynx_hare$hare)
```

```
[1] 47452.38
```

```
>sd(lynx_hare$hare)
```

```
[1] 36885.5
```

Let us now repeat it for lnyx population:

```
>mean(lynx_hare$lynx)
```

```
[1] NA
```

NA is a special character in R and stands for not applicable/not available. Because in some years the data is missing for lynx population, we get the NA result. To ignore the years when the data was missing:

```
>mean(lynx_hare$lynx, na.rm = TRUE)
>sd(lynx_hare$lynx, na.rm = TRUE)
```

```
[1] 2478.327
[1] 1910.491
```