The objective of this section is to review the syllabus and to introduce the R environment. If there is remaining time, I'll work through some basic code puzzles that will require you to work in R, but will more likely leave them for you to play with on your own. Today may be a little slow for those of you with substantial experience in R, but I promise we'll speed up soon.

Download R: The download of R will vary by operating system, but it will begin here in any event:

cran.r-project.org

The online documentation and installer routines are comprehensive. If you are new to R, then it might make sense to use the Mac or Windows distribution, along with the built-in editor to write and evaluate code. `Rstudio` is a popular IDE that provides a somewhat more user-friendly interface than the base R installation. For the tech-oriented, the Linux distribution is very flexible; and I'd use Emacs with the ESS package for editing. If you are interested in using the Linux distribution and are having trouble with the setup, please see me.

I have included links to a few of the many resources on the web that provide gentle introductions to the R language. Those of you who have no experience with R or with programming in general will find it well worth your time to spend a few hours browsing those in your free time. In section, however, I will focus on presenting examples of code piece-by-piece in order to illustrate certain concepts. As always, please interrupt me with questions at any time.

Working in R

In order to download specific packages that are not bundled with the base distribution of R, such as the `foreign` package, you'll enter the following commands to install and load the package:

```
install.packages("foreign")
library(foreign)
```

Once `foreign` is loaded, you'll have access to all of its constituent functions, including `read.csv` which will convert a comma-separated value worksheet (.csv) into a data frame[1]. We will do that now, loading into memory the `auto.csv` into a data frame called `data`.

```
data <- read.csv("auto.csv", header=TRUE)
```

We can read the names from the data set; but they aren't much help.

```
names(data)
```

```
[1] "V1" "V2" "V3"
```

We can replace the column headers with more descriptive variable names.

```
names(data) <- c("price", "mpg", "weight")
```

To get a sense of the data, list the first six observations:

[1] Note that it is also possible to read in `xls`, `dta`, tab-delimited, and many other types of data using similar functions.

```
head(data)
```

```
  price mpg weight
1  4099  22   2930
2  4749  17   3350
3  3799  22   2640
4  4816  20   3250
5  7827  15   4080
6  5788  18   3670
```

With the columns appropriately named, we can refer to particular variables within the data set using the unique indexing in `R`, where data objects tend to be variants of lists and nested lists.

```
head(data$mpg)
```

```
[1] 22 17 22 20 15 18
```

Next week, we'll do more in-depth analysis of this data.

Linear algebra puzzles: These notes will provide a code illustration of the Linear Algebra review in Chapter 1 of the lecture notes. Don't worry if you can't solve these puzzles. Come back to them later, once we have gone over `R` code in more detail. There are many correct ways to solve these puzzles. We will go over a few solutions in section.

1. Let $\mathbf{I}_5$ be a 5×5 identity matrix. Demonstrate that $\mathbf{I}_5$ is symmetric and idempotent using simple functions in `R`.

2. Generate a 2×2 idempotent matrix $\mathbf{X}$, where $\mathbf{X}$ is not the identity matrix. Demonstrate that $\mathbf{X} = \mathbf{XX}$.

3. Generate two random variables, $\mathbf{x}$ and $\mathbf{e}$, of dimension $n = 100$ such that $\mathbf{x}, \mathbf{e} \sim N(0, 1)$. Generate a random variable $\mathbf{y}$ according to the data generating process $y_i = x_i + e_i$. Show that if you regress $\mathbf{y}$ on $\mathbf{x}$ using the canned linear regression routine `lm()`, then you will get an estimate of the intercept β_0 and the coefficient on $\mathbf{x}$, β_1, such that $\beta_0 = 0$ and $\beta_1 = 1$.

4. Show that if $\lambda_1, \lambda_2, \ldots, \lambda_5$ are the eigenvectors of a 5×5 matrix $\mathbf{A}$, then $\text{tr}(\mathbf{A}) = \sum_{i=1}^{5} \lambda_i$.