

Math 214

Applied Statistics

How to get data from TEXT into R

Prof. Verzani has translated almost all of the data sets from the textbook in R readable format. Thank him when you see him. This will make doing the homework easier!

To get the data do the following in an R session:

```
> f = "http://wiener.math.csi.cuny.edu/verzani/classes/Old/S2005/MTH214/datasets/all.R"
> source(url(f))    %% Loads ALL the Kitchens Data sets (almost all)
> ls()              %% Lists the data sets you just loaded. (Lots!)
```

Now, if you SAVE the workspace when you quit (q()) R, these data sets will be available every time you start R.

To access the data, do the typical R stuff, `attach(name.of.dataset)` and `names(name.of.dataset)`.

For example, if I want to look at Example 3.17 on linear regression on page 187 of the text book, I would attach the data set called GPA (Kitchens called it 'Gpa.mtw').

```
> attach(GPA)
> names(GPA)
[1] "HSGPA"  "CollGPA"
> plot(CollGPA~HSGPA)    %% Makes a scatter plot - like Figure 3.18
> cor(CollGPA,HSGPA)    %% How strong is linear relationship?
> res = lm(CollGPA~HSGPA) %% Do linear regression to find best fit line
> res
```

Call:

```
lm(formula = CollGPA ~ HSGPA)
```

Coefficients:

(Intercept)	HSGPA
-0.9504	1.3470

```
>abline(res)    %%Plot the best fit line on the scatter plot (Fig 3.21)
```

NOTE: Most of Kitchens datasets are stored now in R. The naming convention is to change Kitchens' names to ALL CAPS.