4181 - R Practice Questions

Choose one:

- **Three Intermediate Level Questions;**
- **Two Slightly Harder Questions.**

Intermediate Questions

Intermediate 1

- Create four different variables.
    - First one should contain all even integers from 1 to 10.
    - Second one should contain five characters named “DC”, “DE”, “MD”, “PA” and “VA”.
    - Third variable should be the same as the second one, except it should be a `factor`.
    - Fourth variable should be a logical vector that contains 3 `TRUE`s and 2 `FALSE`s.
- What kind of error messages do you expect to see (if any) when you multiply each of those four variables by 2 (if `x` is your variable, `x*2`)? Make a guess and check it in R.

Intermediate 2

- Load the iris data set (just use `data(iris)`), and model the `Sepal.Length` of flowers using simple regression with `Sepal.Width` as your covariate (a.k.a. independent variable, predictor).

Intermediate 3

- Write a simple for loop that calculates the Nth Fibonacci number for a given N (Fibonacci numbers are defined by the recurrence relation:

  $$F(n) = F(n-1) + F(n-2),$$

  with $F(0) = 0, F(1) = 1$.

Intermediate 4

- Find an online dataset and load it into R (this could be a dataset from a package).

Provide some exploratory plots.

Intermediate 5&6 (counts as two questions)

- Create two random vectors with

```
xVec <- sample(0:999, 250, replace=T)
yVec <- sample(0:999, 250, replace=T)
```

- Pick out the values in `yVec` which are > 600.
- What are the index positions in `yVec` of the values which are > 600?
- What proportion of `yVec` are larger than `xVec`? Over repeated samples what would you expect this proportion to converge to?

Slightly Harder Questions (SHQ)

SHQ 1

- Randomly sample 100 points from a ten-dimensional multivariate Gaussian matrix with mean zero and covariance matrix Σ, where

$$\Sigma_{i,j} = 0.9^{(|i-j|)}.$$

That is,

$$\Sigma_{1,1} = 0.9^0 = 1,$$

and that's true for all $\Sigma_{i,i}$.

And, for all other terms,

$$\Sigma_{3,5} = 0.9^{(2)} = 0.81.$$

The terms of $\Sigma_{i,j}$ decay as they move away from the diagonal.

SHQ 2

- Load the `pscl` package, and the `prussian` dataset Using `ggplot2`, plot the number of kicks in each `corp` for each `year`.

SHQ 3

- Load the dataset in `http://staff.elena.aut.ac.nz/Paul-`

`Cowpertwait/ts/global.dat` to R. Create a `ts` variable with a start date of (1856,1) and ending date of (2005,12). Frequency should be in months.

SHQ 4

- Load the `car` package and the `Hartnagel` dataset. Estimate the effect of the fertility rate (`tfr`) on the female indictable-offense conviction rate (`fconvict`) by using multiple linear regression. Include proper model diagnostics.