

The data file:

`http://people.hsc.edu/faculty-staff/blins/StatsExamples/arctic.csv`

contains data about the total amount of water that flows into the Arctic ocean each year from the six largest rivers in Europe and Asia. The discharge is measured in cubic kilometers of water. Using R Studio, create an R markdown document and load the data with the command

```
arcticData = read.csv("arctic.csv")
```

You'll have to save a copy of the data file in the same folder as your R markdown document (or copy the whole URL instead of `arctic.csv`). Here is a list of R commands for working with matrices: <https://www.statmethods.net/advstats/matrix.html>

1. Make a scatterplot of the data.
2. Find the correlation between  $x$  and  $y$  using the `cor()` command and using the formula

$$r = \frac{(x - \bar{x})^T (y - \bar{y})}{\|x - \bar{x}\| \|y - \bar{y}\|}$$

Do you get the same result?

3. Find the matrix  $X$  and the vector  $y$ , and find the least squares solution to  $X\beta = y$ . Hint: You can use the `cbind()` command to combine vectors into columns of a matrix. The command `rep(1,n)` makes a vector with  $n$  copies of the number 1, and the command to multiply matrices in R is `%*%` (the symbol `*` multiplies matrices entrywise instead of the right way).
4. Calculate the standard error of the residuals

$$s = \sqrt{\frac{\|y - \hat{y}\|^2}{n - 2}}$$

5. Use the fact that  $t = \frac{\hat{\beta}_1}{(s/\|x - \bar{x}\|)}$  has a t-distribution with  $n - 2$  degrees of freedom to determine if there is a significant correlation between year and discharge.
6. Make a 95% confidence interval for the slope of the regression line (again, use the fact that  $\frac{\hat{\beta}_1}{(s/\|x - \bar{x}\|)}$  has a t-distribution).