

How to Work

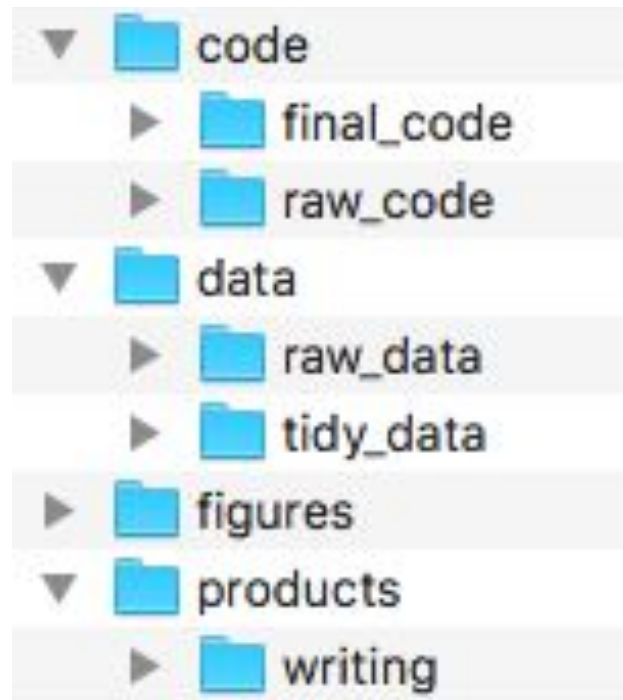


Organizing Data Science Projects

Why organize?

- It makes collaboration easier.
- It reduces the likelihood of making mistakes.
- It makes is a lot easier to go back to your analysis
- It shows transparency





This is the README file for my_first_project

Last updated: 02-Mar-2018

The folders in this project are:

- *data* - is the folder where you can find all the collected data.
- *figures* - is where you can find all the plots, data pictures, and other images.
- *code* - is where you can find code files for collecting, cleaning up, or analyzing data.
- *products* - is where you can find reports, presentations, or products

Data on crime is obtained from International Crime Data collected between 2015-2018 and is publicly available. Data on happiness is collected from the Survey of International Happiness.

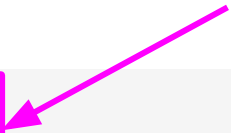
Contributors:

- Jane Everyday Doe, jane.everyday.doe@gmail.com
- John Everyday Doe, john.everyday.doe@gmail.com

Cite: Doe, J, and Doe, J, Sample Analysis Using Sample Data, Working Paper, 2018

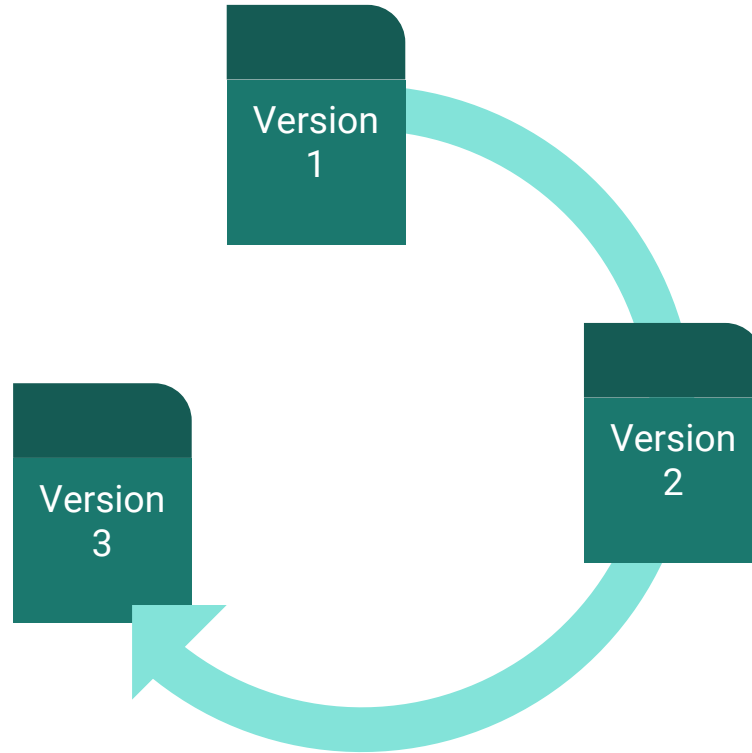
Use Comments

```
# calculates the products of a vector and a matrix  
# checks if they can be multiplied first
```

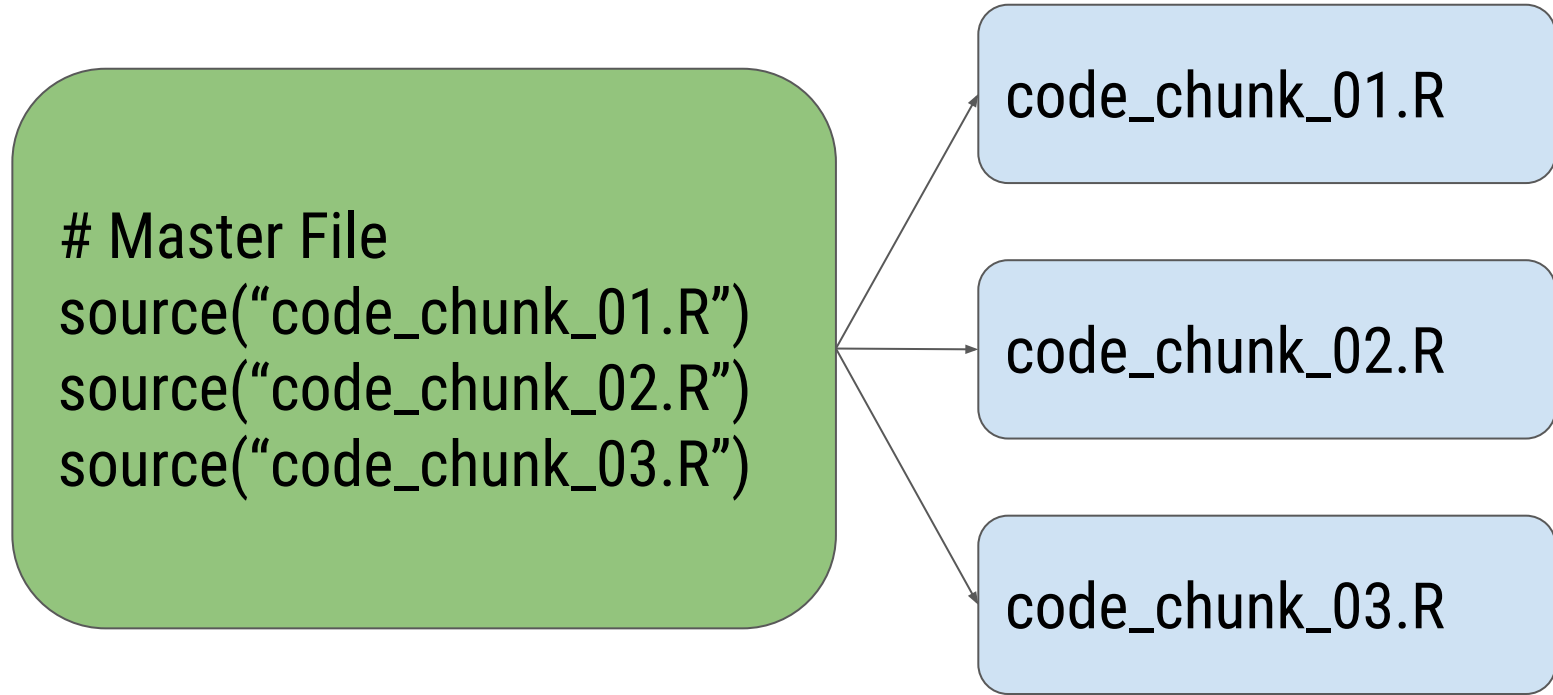


```
function (x, y){  
  if (dim(x)[2] != length(y)) {  
    stop("Can't multiply matrix%*%vector because the dimensions are wrong")  
  }  
  product <- matrix %*% vector  
  return(product)  
}
```

Version Control



Write in a Modular Way



Final Steps

- Use proper file numbering
- Keep updating the README file
- Work in two steps
 - Fast, not worrying about mistakes in raw code
 - Slow, careful to organize and communicate in final
- Write a final report in RMarkdown
- Keep track of what you need to do in a bulleted list
- Keep track of what you do across your projects in a journal

