

SOCIAL DATA SCIENCE

DATA SCIENCE TOOLS

Sebastian Barfort

August 11, 2016

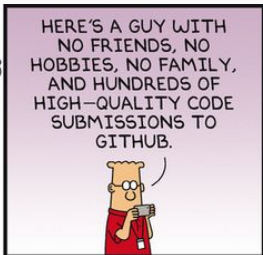
University of Copenhagen
Department of Economics

Git & Github

markdown & rmarkdown



Dilbert.com DilbertCartoonist@gmail.com



4-4-15 © 2015 Scott Adams, Inc./Dist. by Universal Uclick



Git & Github

WHY VERSION CONTROL?

Your closest collaborator is you six months ago, but you don't reply to emails.

Detailed log of all changes.






Easy to revert back to previous versions.

Clear attribution of work (who contributed what).

Without version control

Coordinates_rawdata.csv	18/05/2015 19:07
coords_2015-09-09.csv	23/09/2015 17:18
coords_2015-09-09_modif.csv	05/11/2015 15:20
coords_2015-10-11_modif_YC.csv	17/11/2015 13:49
coords_2015-10-18_modif_YC.csv	18/11/2015 17:26
coords_2015-12-26_modif_YC.csv	28/12/2015 13:33
coords_2015-12-26_modif_YC_years.csv	30/03/2016 19:38
Pulido et al_SMI_Data.csv	20/10/2015 11:55
Pulido et al_SMI_Data_modif_YC_2015-12-26.csv	28/12/2015 13:30
qualitative_data.csv	04/07/2016 15:50
cleandata.xlsx	25/06/2015 01:14
cleandata_YC.xlsx	30/06/2015 16:22
COORDENADAS PACO_20-05-2016 CON REVIEWS.xlsx	20/05/2016 16:23
COORDENADAS PACO_20-05-2016 CON REVIEWS_FRS.xlsx	27/05/2016 19:41
COORDENADAS_paper195(Girella_elevata).xlsx	08/06/2016 13:09
coordenadas_raw_2016-06-08.xlsx	09/06/2016 15:53
coordenadas_raw_2016-06-08_old.xlsx	08/06/2016 16:00
coordenadas_raw_2016-06-21.xlsx	21/06/2016 16:12
coords_2015-09-09_modif.xlsx	05/11/2015 15:23
coords_2015-10-11_modif_YC.xlsx	17/11/2015 13:37
coords_2015-10-11_modif_YC_PACO.xlsx	17/11/2015 17:06
coords_2015-10-18_modif_YC.xlsx	18/11/2015 17:24
coords_2015-12-26_modif_YC.xlsx	30/03/2016 19:38
coords_2016-04-02.xlsx	06/04/2016 17:46
coords_2016-04-02_YC.xlsx	06/04/2016 18:03
coords_2016-04-08_YC.xlsx	11/04/2016 13:51
dataset_y_coords_09_09_15.xlsx	23/09/2015 17:18
Datos metaanalysis_18-04-2016.xlsx	19/04/2016 16:24
FINAL METAANALYSIS_14-6-2016_WITH REVIEWS.xlsx	21/06/2016 16:15
FINAL METAANALYSIS_16-6-2016_WITH REVIEWS.xlsx	21/06/2016 16:13
FINAL METAANALYSIS_2016-04-27_WITH REVIEWS.xlsx	25/05/2016 18:05
FINAL METAANALYSIS_2016-04-27_WITH REVIEWS_FRS.xlsx	27/05/2016 18:44
FINAL METAANALYSIS_2016-04-29_EXCLUDING REVIEWS.xlsx	08/06/2016 13:06
FINAL VOTECOUNTING_1-7-2016.xlsx	04/07/2016 15:46
fitnessdata_2016-06-22.xlsx	22/06/2016 21:00
IFs for Bastien_19-3-2016_YC.xlsx	28/03/2016 19:26
Metaanalysis final_01-05-2015 with coordinates.xlsx	18/05/2015 19:20
Metaanalysis final_22-05-2015 coords.xlsx	24/06/2015 15:50
Metaanalysis final_25-06-2015.xlsx	30/06/2015 16:55
Metaanalysis y coords revisadas_06-08-2015_AH JE.xlsx	23/09/2015 12:57
Pulido et al_SMI_Data_2016-05-27.xlsx	27/05/2016 18:48
Pulido et al_SMI_Data_2016-05-27_PACO.xlsx	08/06/2016 16:33

With version control

 exclosure_damage_raw.csv	04/07/2016 21:21
 exclosures_cover_raw.csv	04/07/2016 20:49
 sitenames.csv	04/07/2016 20:42
 sites_info_raw.csv	30/06/2016 20:03
 species_info_raw.csv	05/07/2016 15:53

what is git? Git is an open source command line program for version control.

what is github? A company/web service that hosts Git repositories and enables 'social coding'

What is GitHub for Mac/Windows? A GUI for Git.

Makes it easier to use.

Ultimately just does command line Git.

Key terms (local):

- Repository (repo): a directory where Git looks for changes
- Initialize (init): have Git begin watching a directory
- add: stage a file so that Git starts watching it
- master: the main branch. By convention this should be the most stable version.
- push: commit changes to a remotely hosted repository
- pull: merge changes from a remotely hosted repository

In GUI GitHub push and pull are combined into sync

markdown & rmarkdown

WHY USE MARKDOWN?

Easy to learn and use.

Focus on **content**, rather than **coding** and debugging **errors**.

It's flexible. Markdown was created to simplify HTML, but with the right tools, your Markdown files can easily be converted to many different formats!

WHAT IS MARKDOWN?

Markdown is a particular type of markup language.

Markup languages are designed produce documents from plain text.

Some of you may be familiar with *LaTeX*. This is another (less human friendly) markup language for creating pdf documents.

LaTeX gives you much greater control, but it is restricted to pdf and has a **much** greater learning curve.

markdown was created for the web (you know it if you use Github, Stackoverflow, etc.)

Suppose we want to create a nested list

- fruits
 - apples
 - macintosh
 - pears
 - peaches
- vegetables
 - chard

```
\begin{itemize}
\item fruits
\begin{itemize}
\item apples
  \begin{itemize}
    \item macintosh
  \end{itemize}
\item pears
\item peaches
\end{itemize}
\item vegetables
  \begin{itemize}
    \item chard
  \end{itemize}
\end{itemize}
```

```
<ul>
  <li>fruits
    <ul>
      <li>apples
        <ul>
          <li>macintosh</li>
        </ul></li>
      <li>pears</li>
      <li>peaches</li>
    </ul></li>
  <li>vegetables
    <ul>
      <li>broccoli</li>
    </ul></li>
</ul>
```

- * fruits
 - apples
 - macintosh
 - pears
 - peaches
- * vegetables
 - chard

WHAT IS `rmarkdown`?

`rmarkdown` is a new package from RStudio that merges a lot of technologies in an exciting way.

In fact, in RStudio now, you integrate R code into your Markdown document and render as HTML/pdf/Word in the click of a button!