Best Practices for the Political Scientist

J. Alexander Branham

Fall 2015

A Quick Overview...

1. Write using plain text

A Quick Overview...

1. Write using plain text
2. Code well

A Quick Overview...

1. Write using plain text
2. Code well
3. Use a version control system

Write using plain text!

- What if we write an article now that gets famous?

Write using plain text!

- What if we write an article now that gets famous?
- 20 years later, some grad student wants to extend our work

Write using plain text!

- What if we write an article now that gets famous?
- 20 years later, some grad student wants to extend our work
- How did we make Figure 1?

Write using plain text!

- What if we write an article now that gets famous?
- 20 years later, some grad student wants to extend our work
- How did we make Figure 1?
- Non-plain text files may be unuseable 20 years from now

Code well - DO NOT CLICK A MOUSE

- All results should be replicable from a source file

Code well - DO NOT CLICK A MOUSE

- All results should be replicable from a source file
- These source files (for example, .R for R scripts or .do for Stat do-files) should be liberally commented

Code well - DO NOT CLICK A MOUSE

- All results should be replicable from a source file
- These source files (for example, .R for R scripts or .do for Stat do-files) should be liberally commented
- Comments explain what you are doing to your future self, collaborators, and others

Comment example

```
# This code creates Fig 1
# I use the mtcars dataset (included with R)
library(ggplot2)
ggplot(mtcars, aes(mpg, wt)) +
  geom_point() +
  geom_smooth(method="lm") # Adds OLS line with SEs
ggsave("fig/fig1.pdf")
```

- Save this code snippit as `fig1.R` (or similar)

Literate programming and reproducible research

- Previous example - have two files for one plot in a paper

Literate programming and reproducible research

- Previous example - have two files for one plot in a paper
  - The paper itself (document.tex or similar)

Literate programming and reproducible research

- Previous example - have two files for one plot in a paper
  - The paper itself (document.tex or similar)
  - the script to create the figure (fig1.R or similar)

Literate programming and reproducible research

- Previous example - have two files for one plot in a paper
  - The paper itself (document.tex or similar)
  - the script to create the figure (fig1.R or similar)
- What if we could combine these to have everything in one easy-to-read file?

Literate programming and reproducible research

- Previous example - have two files for one plot in a paper
  - The paper itself (document.tex or similar)
  - the script to create the figure (fig1.R or similar)
- What if we could combine these to have everything in one easy-to-read file?
  - This is what literate programming is all about!

Literate programming example (using knitr)

```
\begin{section}
This is an example paragraph, writte in \LaTeX.
Using knitr, we can include R code in the following manner
I can reference the figure number by calling ref:
Figure \ref{fig:mpg-and-weight}.
% NTS - updating that figure with squared x doesn't change
\begin{figure}
\centering
<<fig1plot>>=
# I use the mtcars dataset (included with R)
library(ggplot2)
ggplot(mtcars, aes(mpg, wt)) +
  geom_point() +
  geom_smooth(method="lm") # Adds OLS line with SEs
@
\caption{Miles per gallon and weight}
\label{fig:mpg-and-weight}
\end{figure}
```

Version Control

- **YOU NEED TO USE VERSION CONTROL!!!!!**

Version Control

- **YOU NEED TO USE VERSION CONTROL!!!!!**
- This can be as simple as putting things in Dropbox, which enables you to recover previous file versions automatically

Version Control

- **YOU NEED TO USE VERSION CONTROL!!!!!**
- This can be as simple as putting things in Dropbox, which enables you to recover previous file versions automatically
- Word's "track changes" feature...

Pathologies of the Dropbox/Word approach:

- Assumes they're both going to be around (in the same form(ish)) in the future

Pathologies of the Dropbox/Word approach:

- Assumes they're both going to be around (in the same form(ish)) in the future
  - Remember computers 20 years ago???

Pathologies of the Dropbox/Word approach:

- Assumes they're both going to be around (in the same form(ish)) in the future
  - Remember computers 20 years ago???
- You can "clobber" work (save over good work)

Pathologies of the Dropbox/Word approach:

- Assumes they're both going to be around (in the same form(ish)) in the future
  - Remember computers 20 years ago???
- You can "clobber" work (save over good work)
- Must tell coauthors not to touch files while you're working on them

Pathologies of the Dropbox/Word approach:

- Assumes they're both going to be around (in the same form(ish)) in the future
  - Remember computers 20 years ago???
- You can "clobber" work (save over good work)
- Must tell coauthors not to touch files while you're working on them
  - Though this may be changing...

Pathologies of the Dropbox/Word approach:

- Assumes they're both going to be around (in the same form(ish)) in the future
  - Remember computers 20 years ago???
- You can "clobber" work (save over good work)
- Must tell coauthors not to touch files while you're working on them
  - Though this may be changing...
- You also tend to end up with many versions of the same project

Pathologies of the Dropbox/Word approach:

- Assumes they're both going to be around (in the same form(ish)) in the future
  - Remember computers 20 years ago???
- You can "clobber" work (save over good work)
- Must tell coauthors not to touch files while you're working on them
  - Though this may be changing...
- You also tend to end up with many versions of the same project
  - ((show example paper))

Git

- Git is a formal *distributed version control system*

Git

- Git is a formal *distributed version control system*
- It is an easy way to keep track of all the revisions you have saved

Git

- Git is a formal *distributed version control system*
- It is an easy way to keep track of all the revisions you have saved
- It enables easy collaborating between many different people, without the need to email files back and forth or tell them when you're working on something

Git

- Git is a formal *distributed version control system*
- It is an easy way to keep track of all the revisions you have saved
- It enables easy collaborating between many different people, without the need to email files back and forth or tell them when you're working on something
- You only have *one* version of a file at any one time

Git

- Git is a formal *distributed version control system*
- It is an easy way to keep track of all the revisions you have saved
- It enables easy collaborating between many different people, without the need to email files back and forth or tell them when you're working on something
- You only have *one* version of a file at any one time
- You can see the entire history of a file easily

Git

- Git is a formal *distributed version control system*
- It is an easy way to keep track of all the revisions you have saved
- It enables easy collaborating between many different people, without the need to email files back and forth or tell them when you're working on something
- You only have *one* version of a file at any one time
- You can see the entire history of a file easily
- You can see exactly what changed in each new commit

Git

- Git is a formal *distributed version control system*
- It is an easy way to keep track of all the revisions you have saved
- It enables easy collaborating between many different people, without the need to email files back and forth or tell them when you're working on something
- You only have *one* version of a file at any one time
- You can see the entire history of a file easily
- You can see exactly what changed in each new commit
- There are *many* different GUIs so you don't have to deal with the command line

Git

- Git is a formal *distributed version control system*
- It is an easy way to keep track of all the revisions you have saved
- It enables easy collaborating between many different people, without the need to email files back and forth or tell them when you're working on something
- You only have *one* version of a file at any one time
- You can see the entire history of a file easily
- You can see exactly what changed in each new commit
- There are *many* different GUIs so you don't have to deal with the command line
  - I like Sourcetree (for Windows) and SmartGit (for Linux)

Git

- Git is a formal *distributed version control system*
- It is an easy way to keep track of all the revisions you have saved
- It enables easy collaborating between many different people, without the need to email files back and forth or tell them when you're working on something
- You only have *one* version of a file at any one time
- You can see the entire history of a file easily
- You can see exactly what changed in each new commit
- There are *many* different GUIs so you don't have to deal with the command line
  - I like Sourcetree (for Windows) and SmartGit (for Linux)
- ((show example git history))

Github

- Github is the most popular online git service

Github

- Github is the most popular online git service
  - There are many others, such as BitBucket

Github

- Github is the most popular online git service
  - There are many others, such as BitBucket
- Each project gets a repository (“repo”)

Github

- Github is the most popular online git service
  - There are many others, such as BitBucket
- Each project gets a repository (“repo”)
- Each repo is version-controlled (using git)

Github

- Github is the most popular online git service
  - There are many others, such as BitBucket
- Each project gets a repository ("repo")
- Each repo is version-controlled (using git)
- Default is open-source (public)

Github

- Github is the most popular online git service
  - There are many others, such as BitBucket
- Each project gets a repository ("repo")
- Each repo is version-controlled (using git)
- Default is open-source (public)
- You can make repos private (for a fee - students get 5 for free)

Github

- Github is the most popular online git service
  - There are many others, such as BitBucket
- Each project gets a repository ("repo")
- Each repo is version-controlled (using git)
- Default is open-source (public)
- You can make repos private (for a fee - students get 5 for free)
- This file is a part of my "math-camp" repo here