Sharing your work using Markdown and GitHub

Annie Helms

Before this workshop . . .

Create a GitHub profile

https://github.com/

Download Anaconda Navigator

https://www.anaconda.com/products/individual

Download GitBash

https://git-scm.com/downloads

Download R and Rstudio

https://www.r-project.org/

https://rstudio.com/products/rstudio/download/

Click green “Clone or download” button and download ZIP

https://github.com/annie-helms/github_tutorial

Goals for this Workshop

Brief overview of GitHub

Getting started with Jupyter Notebooks

Getting started with Rmarkdown

Creating your first GitHub repository

NOTE: This workshop will only include basic instruction in Markdown, not Python or R code.

What is GitHub?

Online portfolio of your work

Keeps track of version history for each document

Can view and download other people's work

Streamlines collaboration on projects

What can you upload?

Scripts in programming languages that show your analysis for a research project*

*in accordance with consent forms

Challenge projects online (see https://tinyurl.com/y5g5lnwj) to show off coding skills

Any file that is part of your projects (csv, ppt, PDF, Rmd, ipynb, etc.)

Example: https://github.com/annie-helms/barcelona-2019

annie-helms/barcelona-2019: Thi

github.com/annie-helms/barcelona-2019

Apps GitHub Skills Network EZproxy IBM Cloud cat.dicc cat.conj Other bookmarks

Search or jump to... / Pull requests Issues Marketplace Explore

annie-helms / barcelona-2019

Unwatch 1 Star 0 Fork 0

Code Issues 0 Pull requests 0 Actions Projects 0 Wiki Security 0 Insights Settings

This repository contains the analysis for my sociophonetic research in Barcelona from the summer of 2019.

Manage topics

10 commits 1 branch 0 packages 0 releases 1 co

Branch: master New pull request Create new file Upload files Find file Clo

annie-helms Update README.md Latest commit 0835943 3 days ago

README.md	Update README.md	3 days ago
barcelona_vowels.Rmd	fixed Markdown syntax	3 days ago
barcelona_vowels.csv	data sheet with formant productions	3 days ago
barcelona_vowels.html	updated with k-means clustering	3 days ago
blp_scores.csv	blp_scores.csv	3 days ago

Brief description of repository contents.

Type here to search

ENG US 9:38 AM 6/5/2020

annie-helms/barcelona-2019: Thi

github.com/annie-helms/barcelona-2019

Apps GitHub Skills Network EZproxy IBM Cloud cat.dicc cat.conj Other bookmarks

Search or jump to... / Pull requests Issues Marketplace Explore

annie-helms / barcelona-2019

Unwatch 1 Star 0 Fork 0

Code Issues 0 Pull requests 0 Actions Projects 0 Wiki Security 0 Insights Settings

This repository contains the analysis for my sociophonetic research in Barcelona from the summer of 2019.

Edit

Manage

Files in repository

1 branch 0 packages 0 releases 1 contributor

Branch: master ▾ ew pull request

Create new file Upload files Find file Clone or download ▾

annie-helms e README.md Latest commit 0835943 3 days ago

README.md	Update README.md	3 days ago
barcelona_vowels.Rmd	fixed Markdown syntax	3 days ago
barcelona_vowels.csv	data sheet with formant productions	3 days ago
barcelona_vowels.html	updated with k-means clustering	3 days ago
blp_scores.csv	blp_scores.csv	3 days ago

Type here to search

ENG US 9:38 AM 6/5/2020

annie-helms/barcelona-2019: Thi

github.com/annie-helms/barcelona-2019

Apps GitHub Skills Network EZproxy IBM Cloud cat.dicc cat.conj Other bookmarks

Search or jump to... / Pull requests Issues Marketplace Explore

annie-helms / barcelona-2019

Unwatch 1 Star 0 Fork 0

Code Issues 0 Pull requests 0 Actio Security 0 Insights Settings

This repository contains the analysis for my sociophone the summer of 2019.

Edit

Manage topics

10 commits 1 branch 0 releases 1 contributor

Comments explaining what was updated in this version

Branch: master New pull request Create new file Upload files Find file Clone or download

annie-helms Update README.md Latest commit 0835943 3 days ago

README.md	Update README.md	3 days ago
barcelona_vowels.Rmd	fixed Markdown syntax	3 days ago
barcelona_vowels.csv	data sheet with formant productions	3 days ago
barcelona_vowels.html	updated with k-means clustering	3 days ago
blp_scores.csv	blp_scores.csv	3 days ago

Type here to search

ENG US 9:38 AM 6/5/2020

annie-helms/barcelona-2019: Thi

github.com/annie-helms/barcelona-2019

Apps GitHub Skills Network EZproxy IBM Cloud cat.dicc cat.conj Other bookmarks

Search or jump to... / Pull requests Issues Marketplace Explore

annie-helms / barcelona-2019

Unwatch 1

Code Issues 0 Pull requests 0 Actions Projects 0 Wiki Security 0 Insights

This repository contains the analysis for my sociophonetic research in Barcelona from the summer of 2019.

Manage topics

10 commits 1 branch 0 packages 0 releases 1 contribu

Branch: master New pull request Create new file Upload files Find file Clone or download

annie-helms Update README.md

README.md	Update README.md	
barcelona_vowels.Rmd	fixed Markdown syntax	
barcelona_vowels.csv	data sheet with formant productions	
barcelona_vowels.html	updated with k-means clustering	3 days ago
blp_scores.csv	blp_scores.csv	3 days ago

Clone with HTTPS Use SSH

Use Git or checkout with SVN using the web URL.

https://github.com/annie-helms/barcelona

Open in Desktop Download ZIP

Allows for downloading of repository contents

Type here to search

ENG US 9:43 AM 6/5/2020

barcelona_vowels.Rmd	fixed Markdown syntax	3 days ago
barcelona_vowels.csv	data sheet with formant productions	3 days ago
barcelona_vowels.html	stering	3 days ago
blp_scores.csv		3 days ago
blp_scores.ipynb		3 days ago
clusters.csv		3 days ago

Scrolling down, this README file explains how the files function together in the project

README.md

barcelona-2019

This repository contains the analysis for my sociophonetic research in Barcelona from the summer of 2019.

The blp_scores.ipynb applies k-mean clustering to participants' responses in the Bilingual Language Profile (BLP) survey. In the model, participants are clustered into three groups, where one group is more balanced, one group favors Catalan, and one group favors Spanish.

In barcelona_vowels.Rmd, this dataset is merged with the dataset containing participants' productions in Spanish and Catalan. The resulting data is analyzed and visualized.

barcelona-2019/blp_scores.ipynb

github.com/annie-helms/barcelona-2019/blob/master/blp_scores.ipynb

Apps | GitHub | Skills Network | IBM Cloud | cat.dicc | cat.conj | Other bookmarks

Search or jump to... / Pull requests Issues Marketplace Explore

annie-helms / barcelona-2019

Unwatch 1 | Star 0 | Fork 0

Code | Issues 0 | Pull requests 0 | Actions | Projects 0 | Wiki | Security 0 | Insights | Settings

Clicking on a file opens a preview page, where you can view older versions and view the file in its entirety

Branch: master ▾ barcelona-2019 / blp_scores.ipynb — Find file | Copy path

annie-helms updated with analysis — cdf0687 3 days ago

1 contributor

1844 lines (1844 sloc) | 63 KB — Raw | Blame | History

Bilingual Language Profile Analysis

In this notebook, I will use K-means clustering to group my participants according to their raw responses to the Bilingual Language Profile (BLP) survey (Birdsong, D., Gertken, L.M., & Amengual, M. Bilingual Language Profile: An Easy-to-Use Instrument to Assess Bilingualism. COERLL, University of Texas at Austin. Web. 20 Jan. 2012. https://sites.la.utexas.edu/bilingual/.)

The survey gathers information from two separate languages about the following four categories:

- Attitudes towards the language
- History with the language
- Proficiency in the language
- Use of the language

The survey, as designed by Birdsong et al., weights each category and then computes an overall dominance score for each language. Rather than use the weights given by the authors, I will use the raw responses in each category to group the participants according to their similarity with each other. This categorical output will be computed with three levels, and will hopefully group participants as more Spanish dominant, as more Catalan dominant, or mostly balanced.

In [12]:
```
import pandas as pd
import numpy as np # handle data as in a vectorized manner

# Matplotlib and associated plotting modules
import matplotlib.cm as cm
import matplotlib.colors as colors
```

Type here to search

ENG US 9:46 AM 6/5/2020

barcelona-2019/blp_scores.ipynb

github.com/annie-helms/barcelona-2019/blob/master/blp_scores.ipynb

Apps GitHub Skills Network IBM Cloud cat.dicc cat.conj Other bookmarks

In this notebook, I will use K-means clustering to group my participants according to their raw responses to the Bilingual Language Profile (BLP) survey (Birdsong, D., Gertken, L.M., & Amengual, M. Bilingual Language Profile: An Easy-to-Use Instrument to Assess Bilingualism. COERLL, University of Texas at Austin. Web. 20 Jan. 2012. https://sites.la.utexas.edu/bilingual/.)

The survey gathers information from two separate languages about the following four categories:

- Attitudes towards the language
- History with the language
- Proficiency in the language
- Use of the language

The survey, as designed by Birdsong et al., weights each category and then computes an overall dominance score for each language. Rather than use the weights given by the authors, I will use the raw responses in each category to group the participants according to their similarity with each other. This categorical output will be computed with three levels, and will hopefully group participants as more Spanish dominant, as more Catalan dominant, or mostly balanced.

In [12]:
```
import pandas as pd
import numpy as np # handle data as in a vectorized manner

# Matplotlib and associated plotting modules
import matplotlib.cm as cm
import matplotlib.colors as colors

# import k-means from clustering stage
from sklearn.cluster import KMeans
```

Input

In [17]:
```
blp = pd.read_csv("blp_scores.csv")
blp.head()
```

Output

Out[17]:

	Participant	Castellano	Catalan	Attitude_Castellano	Attitude_Catalan	History_Castellano	History_Catalan	Proficiency
0	p1	151.914	188.424	20	24	86	96	23
1	p2	126.848	200.958	20	24	87	97	18
2	p3	143.014	200.410	20	24	96	115	19
3	p4	125.674	172.170	12	12	81	120	19
4	p5	175.166	164.446	22	23	104	84	19

For the present analysis, I will drop some columns from the DataFrame, so that only the participant number and their raw scores for each category (4 in each language, 8 total) remain.

In [18]:
```
blp = blp[['Participant', 'Attitude_Castellano', 'Attitude_Catalan', 'History_Castellano', 'History_Catalan', 'Proficiency_Castellano', 'Proficiency_Catalan', 'Use_Castellano', 'Use_Catalan']]
blp.head()
```

What kind of file is .ipynb?

This is a Jupyter Notebook, which contains chunks of code along with their outputs

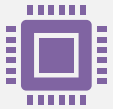
Type here to search

ENG US 9:50 AM 6/5/2020

Jupyter Notebooks

Web application used to create and share code

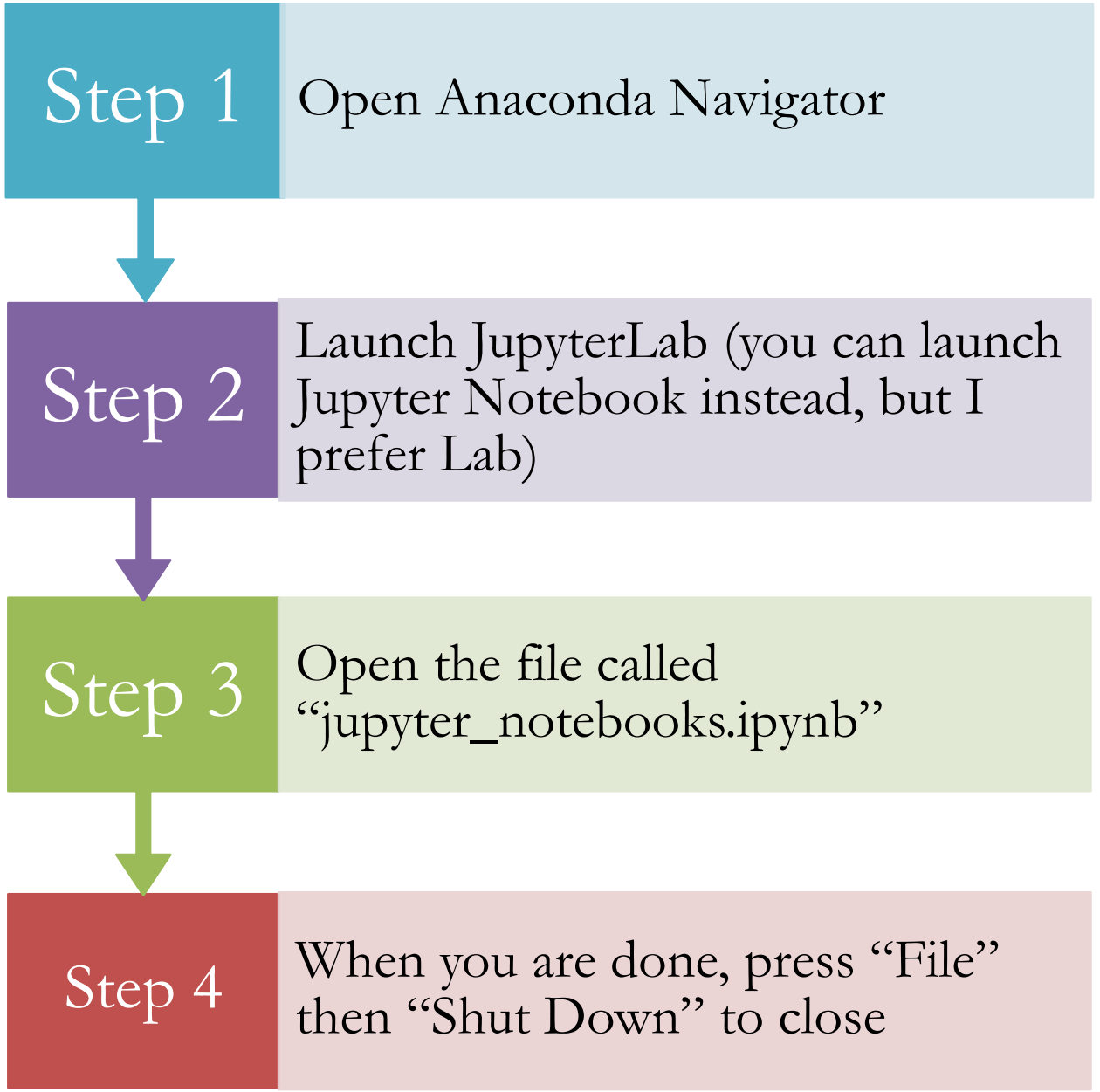
Interactive with code outputs presented in the document



Supports multiple programming languages (R, Python, and more)

Files can be converted to multiple file types (html, PDF, LaTeX, etc.)

Using Jupyter Notebooks



RStudio

User-friendly environment for R

Saves code chunks allowing for collaboration

Includes file type “Rmd” which uses R code and Markdown syntax

Using RMarkdown

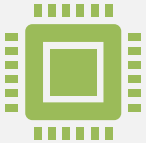
Step 1 Open RStudio

Step 2 Open the file "r_markdown.Rmd" from the GitHub repository

Step 3 When you have finished, click "Knit" to knit file to HTML (or PDF)

GitBash

Syncs your files with your GitHub repository



Authenticates you and your computer using an SSH key

Clones GitHub repositories to your local folder

Creating and Syncing a GitHub Repository

Step 1

Launch GitBash on your computer

Step 2

Enter the following line of code using the email address you used to create a GitHub account:

ssh-keygen -t rsa -b 4096 -C "*email*"

Step 3

Save the SSH key to a folder in your computer (easiest just to press "enter" and accept default location)

Step 4

Choose a password (easiest just to press "enter" and decline to use a password)

Creating and Syncing a GitHub Repository

Step 5

Add the SSH key to the ssh-agent by running the following line of code:

eval "$(ssh-agent -s)"

and then run: **ssh-add ~/.ssh/id_rsa**

Step 6

Run the following line of code to copy your SSH key to your clipboard:

cat ~/.ssh/id_rsa.pub | clip

Step 7

In your GitHub account, click your profile picture, then "Settings"

Step 8

Under "Personal Settings" select "SSH and GPG keys"

Creating and Syncing a GitHub Repository

Step 9	Click “New SSH Key” to add your key to your account
Step 10	Add a title like “personal-computer” and paste your SSH key into the “Key” field
Step 11	Click “Add SSH Key”
Step 12	Create a Github repository by clicking “New” on the left side of the homepage

Creating and Syncing a GitHub Repository

Step 13

Make repository public, initialize with README, and click Create

Step 14

Click green "Clone or Download" button and clone with SSH. Copy the URL

Step 15

Go to Gitbash and run this command to create your folder, using the directory of your choice:

mkdir ~/Desktop/github

Step 16

Navigate to this folder by running:

cd ~/Desktop/github

Creating and Syncing a GitHub Repository

Step 17	Now run the following, pasting in your URL: **git clone *repository_URL***
Step 18	You should see a new folder named after your repository if you run: **ls**
Step 19	Copy the Jupyter notebook and Rmarkdown file you created earlier into your new folder
Step 20	In GitBash, navigate to your folder with: **cd ~/Desktop/github/*repo_name***

Creating and Syncing a GitHub Repository

Step 21	Run the following code to view what changes have been made in your local folder: **git status**
Step 22	To select a file you want to update on GitHub, run: **git add *name_of_file***
Step 23	Commit this change by running: **git commit -m *'summarize changes'***
Step 24	To sync all committed files to GitHub, run: **git push**

Summary of Workshop

Brief overview of GitHub

Getting started with Jupyter notebooks

Getting started with RMarkdown

Creating and syncing a GitHub repository with GitBash