



Laboratorio Nacional de Políticas Públicas



CENTRO DE INVESTIGACIÓN
Y DOCENCIA ECONÓMICAS A.C.

Introducción a Github

Periodismo de Datos
Mayo, 2021

M.C. JORGE JUVENAL CAMPOS FERREIRA.

Investigador Asociado.

Laboratorio Nacional de Políticas Públicas
CIDE



Github es una plataforma de almacenamiento de código para el control de versiones y para la colaboración entre equipos de programadores. Te permite a ti y a otros colaboradores trabajar en proyectos desde cualquier lugar del mundo.



Retomando el tema de la replicabilidad, el poner a disponibilidad del público los códigos con los que trabajamos permiten a los lectores verificar nuestros métodos, aportar información nueva y aprender de nuestro trabajo.

Sin embargo, a pesar de subir las cosas a Github, los trabajos pueden no ser reproducibles.

Que algo se encuentre en Github no es garantía de reproducibilidad o transparencia, pero abona bastante.



Lo relevante de Github:

- **Almacenamiento** de tu código en línea.
- **Colaboración** en equipos o con colaboradores externos.
- **Control de versiones** para respaldar tu trabajo.
- **Red Social** para contactar a otros colaboradores, programadores o personas.
- **Portafolio de trabajo** para mostrar tus capacidades con algún tipo de herramienta.
- **Host gratuito** de páginas web sencillas.
- Una gran fuente de código para **aprender** a programar.



Una red social/profesional para programadores/desarrolladores/analistas.

- Las empresas que contratan perfiles de desarrollador/programador se fijan en que tengas un perfil interesante.
- Puedes seguir el trabajo de otros programadores, y estos pueden seguir también tu trabajo.
- Puedes usar tu perfil como “Portafolio” de tu trabajo como programador y analista, para demostrar de manera pública tus conocimientos y capacidades de código.



Un sistema para compartir código.

- Te permite compartir código con la comunidad. El compartir tu código/ análisis permite que este sea analizado y revisado.
- Te permite ver cuales de tus proyectos despiertan más interés en la comunidad.
- Te permite replicar código (hacer fork) o descargar proyectos/ejemplos/tutoriales/material de otros perfiles.
- Es una fuente muy grande de ejemplos para trabajar.
- Te permite tener un sistema de control de versiones.
- Puedes mostrar a los usuarios las contribuciones y los periodos de trabajo que manejas.
- Puedes compartir proyectos de trabajo personales.

Open Journalism, por Github



Explore Topics Trending Collections Events GitHub Sponsors

[Get email updates](#)

COLLECTION

Open journalism

See how publications and data-driven journalists use open source to power their newsroom and ensure information is reported fairly and accurately.

[Suggest edits](#)

 [fivethirtyeight / data](#) [Star](#)

Data and code behind the articles and graphics at FiveThirtyEight

☆ 14749 ⚡ 9445 Jupyter Notebook

 [datadesk / notebooks](#) [Star](#)

All of our computational notebooks

☆ 311 ⚡ 25 Python

 [nytimes / objective-c-style-guide](#) [Star](#)

The Objective-C Style Guide used by The New York Times

Medios que utilizan Github



NBC News Digital

We are a passionate, creative and nimble team within a legendary news organization, with the power and resources to shape the future of digital news.

⌚ 30 Rockefeller Plz, New York, NY 10020 ⌐ <http://www.nbcnews.com> Verified



Vox Media

⌚ Washington, DC ⌐ <http://product.voxmedia.com>



ProPublica

Journalism in the Public Interest

⌚ New York ⌐ <https://www.propublica.org/> Verified



Time Labs

Greatest Hits From Time Labs

⌚ Washington, D.C. ⌐ <http://time.com> ⏤ chris.wilson@time.com



BBC

Open source code used on public facing services, internal services and educational resources.

⌚ London ⌐ <http://www.bbc.co.uk/opensource/> ⏤ appadmin@bbc.co.uk



Los Angeles Times Data and Graphics Department

Reporting, editing, computer programming.

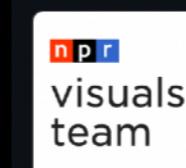
⌚ Southern California ⌐ <https://www.latimes.com> ⏤ @datographics ⏤ datographics@caltimes.com



The New York Times

⌚ New York, NY ⌐ <https://nytimes.com>

Verified



NPR News Apps team

⌚ Washington, DC ⌐ <http://blog.apps.npr.org> ⏤ nprapps@npr.org

```
$ fab -f bees.py attack:h  
Read 2 bees from the roost.  
Connecting to the hive.  
Assembling bees.  
Each of 2 bees will fire 3  
Stinging URL so it will be  
[localhost] run: curl https://  
Organizing the swarm.  
bee 0 is joining the swarm.  
bee 1 is joining the swarm.  
bee 0 is firing his machine.
```

Tribune News Applications Team

⌚ Chicago, IL ⌐ <http://apps.chicagotribune.com>



Bloomberg Media



The Guardian

The source code of the world's leading liberal voice

⌚ London ⌐ <https://www.theguardian.com/> ⏤ @gdndevelopers



FiveThirtyEight

⌚ <https://fivethirtyeight.com> ⏤ dhrumil.mehta@fivethirtyeight.com

Cómo utilizar Github



Ejemplo tomado de la presentación de @maxharlow. Ver [enlace](#).

A screenshot of a BBC News article titled "Five patients made 8,303 emergency calls in a year". The article is from the UK section, specifically England. It features a photo of a busy control room with multiple computer monitors displaying data. The BBC navigation bar at the top includes links for Home, Video, World, US & Canada, UK, Business, Tech, Science, Stories, Entertainment & Arts, Health, and More. A search bar and a GitHub logo are also visible in the top right corner. To the right of the main article, there is a sidebar titled "Top Stories" with three news items and a "Features" section with one item.

Generar un análisis



Subir código al
repositorio



Ponerlo a
disposición del
público

Cómo utilizar Github



Ejemplo tomado de la presentación de @maxharlow. Ver [enlace](#).

The screenshot shows a GitHub repository page for `BBC-Data-Unit/emergency-frequent-callers`. The repository has 9 commits, 1 branch, 0 releases, and 1 contributor. The latest commit was made a day ago. The repository contains files like `Frequent Callers Data.xlsx`, `README.md`, and `average_job_time.csv`. A summary statement at the bottom reads: "Five patients made 8,303 emergency calls in a year".

Generar un análisis

Subir código al
repositorio

Ponerlo a
disposición del
público

Cómo utilizar Github



Ejemplo tomado de la presentación de @maxharlow. Ver [enlace](#).

The screenshot shows a news article from BuzzFeed News. The title is "BuzzFeed News Trained A Computer To Search For Hidden Spy Planes. This Is What We Found." Below the title is a subtitle: "From planes tracking drug traffickers to those testing new spying technology, US airspace is buzzing with surveillance aircraft operated for law enforcement and the military." The author is Peter Aldhous, a BuzzFeed News Reporter. There are social sharing icons for Facebook, Twitter, Email, Pinterest, Telegram, and LinkedIn. The text discusses secret spy planes used by US Marshals and military contractors. It mentions that these planes hunt drug cartel kingpins in Mexico and track terrorists in Africa. The article also notes that these discoveries were made by training a computer to recognize known spy planes using flight-tracking data.

Generar un análisis



Subir código al
repositorio



Ponerlo a
disposición del
público

Cómo utilizar Github



Ejemplo tomado de la presentación de @maxharlow. Ver [enlace](#).

The screenshot shows a GitHub repository page. At the top, there's a navigation bar with links for 'This repository', 'Search', 'Pull requests', 'Issues', 'Marketplace', and 'Explore'. Below the navigation, the repository name 'BuzzFeedNews / 2017-08-spy-plane-finder' is displayed, along with metrics: 11 commits, 2 branches, 0 releases, and 1 contributor. A red horizontal bar highlights the '11 commits' section. Below this, there's a list of files and their commit history:

File	Commit Message	Latest Commit
paldhous more notes on model evaluation	more notes on model evaluation	5d6cdc9 on 10 Aug 2017
data	more notes on model evaluation	7 months ago
README.md	publication date	7 months ago
index.Rmd	more notes on model evaluation	7 months ago
index.html	more notes on model evaluation	7 months ago

Below the file list, a large text block reads: 'BuzzFeed News Trained A Computer To Search For Hidden Spy Planes. This Is What We Found.' A note below it says: 'R code to reproduce the analysis underlying this August 7, 2017 BuzzFeed News article on using machine learning to identify candidate surveillance aircraft from flight tracking data.'

Generar un análisis



Subir código al repositorio



Ponerlo a disposición del público

Alternativas a Github



Alternativas a Github.

- git puro.
- Gitlab
- BitBucket

Tienen funciones muy similares a las que ofrece Github.





Control de versiones.

Un **control de versiones** es un sistema que registra los cambios realizados en un archivo o conjunto de archivos a lo largo del tiempo, de modo que puedas recuperar versiones específicas más adelante.

Atrás/Adelante	Nombre	Visualización	Grupo	Compartir	Agregar etiquetas	Acción	Buscar
	Copia de tesis						
	Copia de tesis 2						
	thesis						
	Tesis Versión Definitiva	hoy 19:03					
	Tesis Versión Final	hoy 19:03					
	Tesis Versión Final Final	hoy 19:03					
	Tesis Versión Final Final_2	hoy 19:03					
	Tesis_v2	hoy 19:03					
	Tesis_v3	hoy 19:03					
	Tesis_v4	hoy 19:03					



Repositorio de Github

Un **repositorio** es la carpeta de tu proyecto, en la que se guardan tus archivos (código, datos, imágenes), el historial de revisión de cada archivo y la administración y el debate propiciado por tu código.

Los repositorios pueden ser públicos y privados (acceso restringido). Igualmente, los repositorios pueden pertenecer a una persona o pueden tener colaboradores asociados trabajando en un mismo proyecto.

A screenshot of a GitHub repository page. The repository name is 'Inpp / metodologia_diagnostico_cuidados_mex'. The 'Code' tab is selected. The main commit is by 'JuveCampos' titled 'Add files via upload' with 22 commits. Below it is a list of other commits, all of which have been merged and are 4 months ago or 29 days ago. The commits are: 'Derecho a cuidar', 'Derecho a ser cuidadx - Accesibilidad', 'Derecho a ser cuidadx - Asequibilidad', 'Derecho a ser cuidadx - Cobertura', 'Derecho a ser cuidadx - Pertinencia', 'Documentos', 'Estimaciones de Crecimiento Poblacional', and 'Multimedia'.

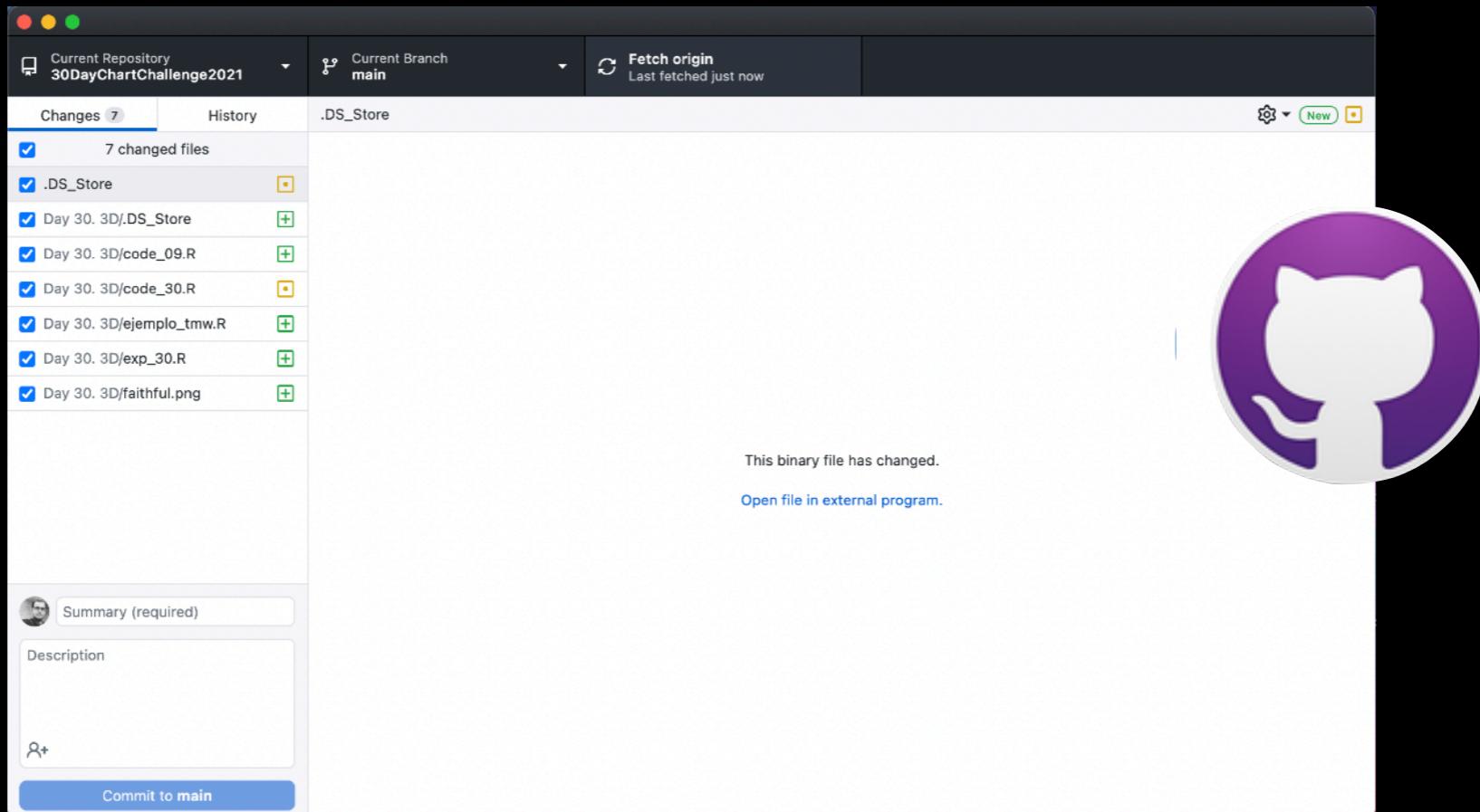
Commit	Author	Date
Derecho a cuidar	JuveCampos	4 months ago
Derecho a ser cuidadx - Accesibilidad	JuveCampos	4 months ago
Derecho a ser cuidadx - Asequibilidad	JuveCampos	4 months ago
Derecho a ser cuidadx - Cobertura	JuveCampos	4 months ago
Derecho a ser cuidadx - Pertinencia	JuveCampos	4 months ago
Documentos	Add files via upload	29 days ago
Estimaciones de Crecimiento Poblacional	addModsPagina1	4 months ago
Multimedia	mods_img_tarjetas	29 days ago

Conceptos clave



Github Desktop.

Github Desktop es un programa que permite al usuario interactuar con Github directamente desde el escritorio sin entrar a la página. Esto reduce algunos pasos de la gestión de archivos, y permite guardar copias locales de nuestros repositorios de manera más sencilla.



Conceptos clave



readme.md

El archivo readme.md es la portada de nuestro repositorio. En este archivo se trata de caracterizar el repositorio, dar instrucciones para su mejor uso o manejo, incluir ejemplos de uso, dar créditos y definir las licencias de uso de la información. (Por ejemplo, licencias MIT o licencia BSD o algún otro tipo de licencia.).

Esta sección se edita en Markdown.

README.md

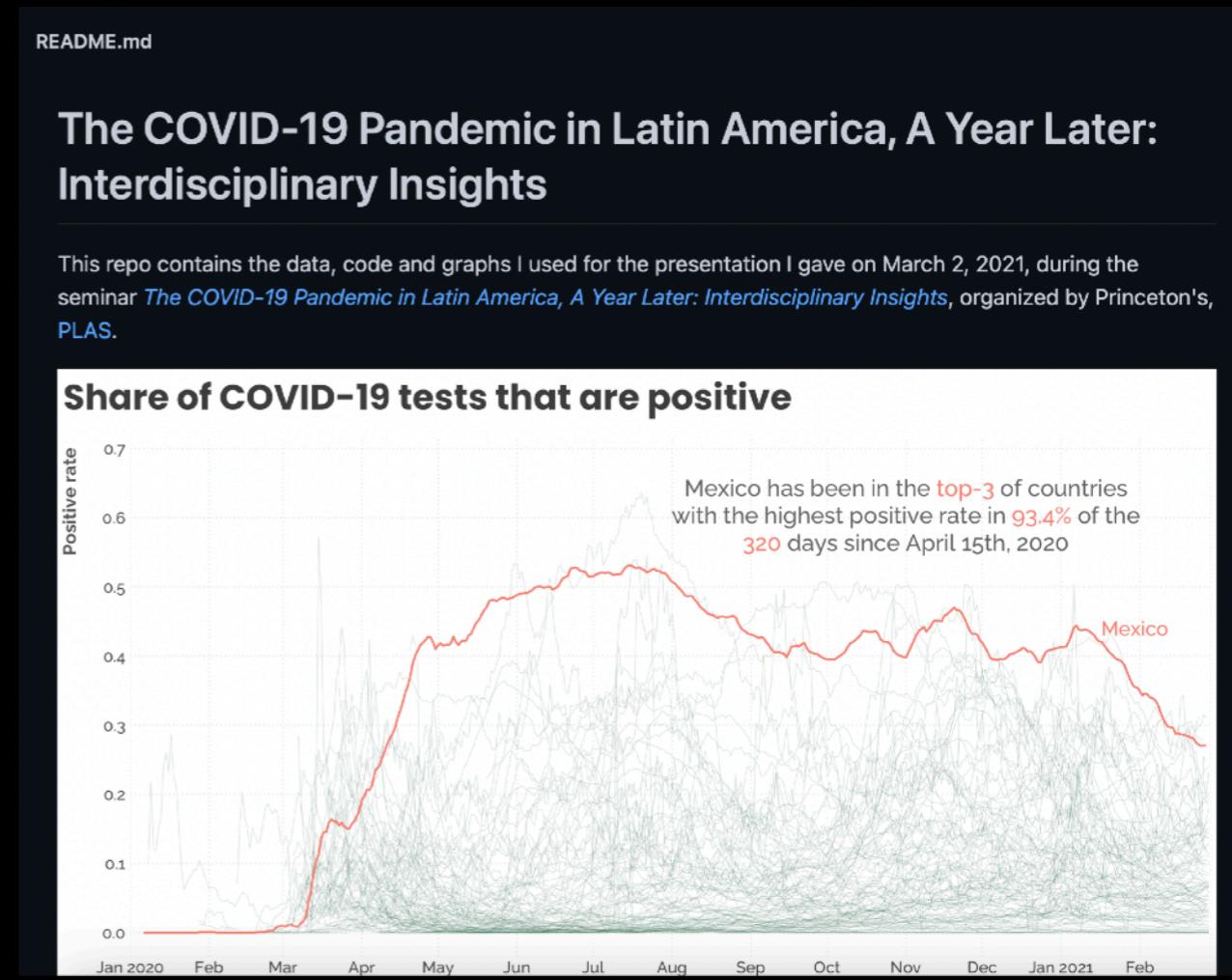
LAS BRECHAS DE CUIDADOS EN LOS MUNICIPIOS DE MÉXICO

Proyecto de metodología, diagnóstico y propuesta de cuidados a nivel municipal

Laboratorio Nacional de Políticas Públicas (LNPP)
Centro de Investigación y Docencia Económicas (CIDE)
15 de enero de 2021

Contenido del repositorio:

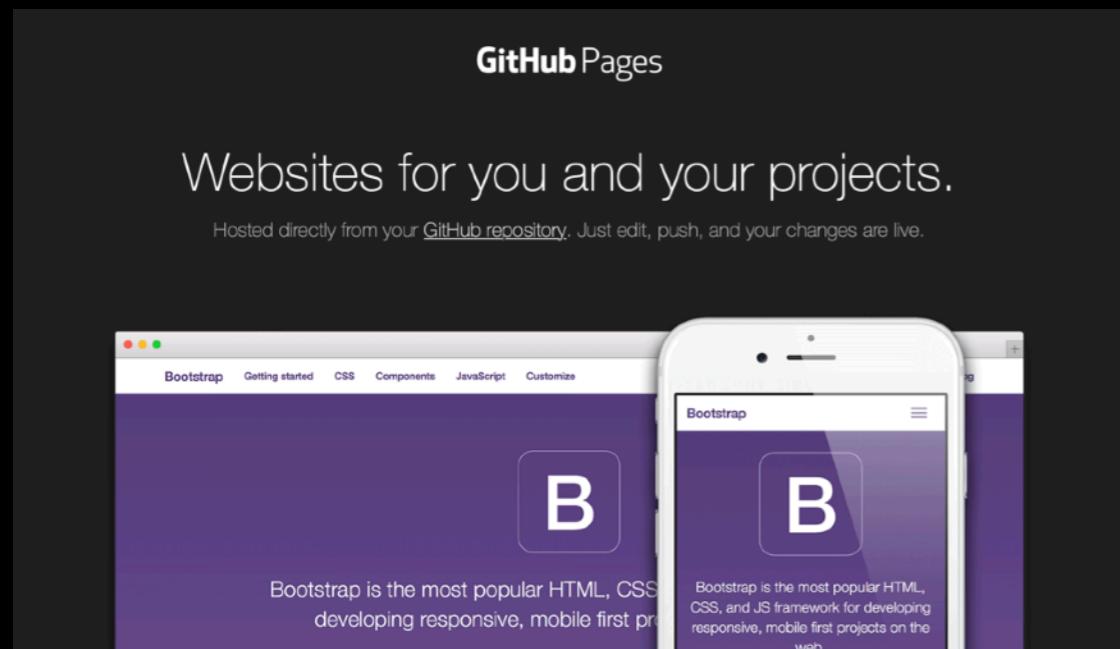
Los scripts contenidos en este repositorio miden los indicadores de los atributos del derecho a cuidar y del derecho a ser cuidada o cuidado, de acuerdo con la metodología desarrollada en el documento "[Marco analítico y metodología para diagnosticar las brechas en los cuidados en municipios de México](#)".





Github Pages.

El servicio de **Github Pages** es un servicio que permite almacenar páginas web estáticas a partir de los archivos presentes en un repositorio. Es una buena primera opción si queremos probar alguna página web sin hacer el gasto en comprar dominio y servidor.



The screenshot shows a personal blog website. At the top, there is a navigation bar with links to "Sobre mí", "GitHub", "Twitter", and "LinkedIn". The main header features a photo of a man and the name "Juve Campos" in red script. Below the header is a photo of a building with large windows. A pink banner across the photo reads "Blog personal 😎👨‍💻". The main content area has a dark background with white text. It features a section titled "Presentaciones" with a sub-section for "2021 #30DayChartChallenge 2". To the right, there is a sidebar with a list of sessions for "Clase de Periodismo de Datos 2021 - MPPP".

- Sesión 1. [Introducción a RMarkdown](#)
- Sesión 3. [Expresiones regulares con rebus](#)
- Sesión 4. [Mapas en ggplot2](#)
- Sesión 5. [Mapas interactivos con leaflet](#)
- Sesión 6. [Visualización interactiva en R](#)
- Sesión 7. [Datos Abiertos y Descargas Automatizadas](#)
- Sesión 8 y 9. [Descarga de datos de APIs](#)
- Sesión 10. [Web Scraping](#)

Conceptos clave



Gists

Si no quieres guardar código a través de un repositorio (porque consideras que un repositorio es demasiado), Github permite almacenar pequeños cachitos de código (snippets) a través de gists.

JuveCampos / block16. Secret
Last active 16 months ago

```
1 # Graficamos
2 ggplot(pop, aes(x = `Grupos quinquenales de edad`,
3                   y = `Poblacion por Sexo`,
4                   fill = Sexo)) +
5   geom_linerange(data = subset(pop, Sexo == "Hombres") %>%
6     # Convertimos los datos de los Hombres en negativos
7     mutate(`Poblacion por Sexo` = -`Poblacion por Sexo`),
8     aes(ymin = -desplazamiento,
9          ymax = -desplazamiento +`Poblacion por Sexo`),
10    size = 5,
```

JuveCampos / block15. Secret
Created 16 months ago

```
1 {r}
2 # LABELS - DEFINEN EL CONTENIDO
3 c(rev(seq(0, 180000, by = 30000)),
4   seq(0, 180000, by = 30000))
```

JuveCampos / block14. Secret
Created 16 months ago

```
1 {r}
2 # BREAKS - DEFINEN LA POSICION
3 c(seq(-180000, 0, by = 30000) - desplazamiento,
4   seq(0, 180000, by = 30000) + desplazamiento)
```

jjsantos01 / dgisConexion.ipynb
Last active 9 days ago

Pasos para conectarse a los cubos dinámicos de la DGIS de la Secretaría de Salud

Es necesario instalar [adodbapi](#)

```
pip install adodbapi
```

In [2]:
import pandas as pd
import adodbapi

- Se necesita descargar el driver de MSOLAP (amd64): <https://docs.microsoft.com/en-us/analysis-services/client-libraries?view=asallproducts-allversions&viewFallbackFrom=sql-server-2017>
- Probablemente también haya que instalar los drivers que especifica la páginas de DGIS: http://www.dgis.salud.gob.mx/contenidos/basesdedatos/BD_Cubos_gobmx.html

jjsantos01 / parejas.ipynb
Created 9 days ago

Calcula hogares con parejas homosexuales a partir del censo 2020

In [178]:
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
import glob

<https://www.animalpolitico.com/capital-plural/las-parejas-del-mismo-sexo-en-mexico-cuentan/>
<https://www.animalpolitico.com/blog-invitado/las-familias-homoparentales-ya-existen-no-se-crean-con-leyes>

Gists de JuveCampos.

Jorge Juvenal Campos Ferreira

Gists de jjsantos01.

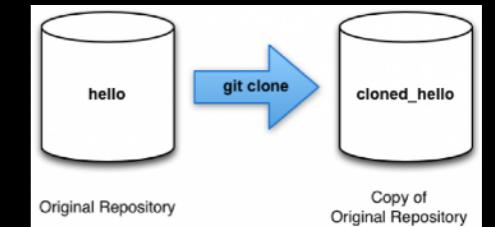
MPPP – PdD2020



Colaboración en Github.

Para colaborar en github, hay que saber algunos de los términos necesarios para entender que acciones estamos llevando a cabo.

Clone



Clone, se utiliza principalmente para clonar o copiar un repositorio remoto en un nuevo directorio, en otra ubicación (generalmente en nuestra computadora, creando una copia local).

Pull

1 ↓

El comando pull se emplea para **extraer y descargar contenido desde un repositorio remoto y actualizar al instante el repositorio local para reflejar ese contenido**. Hay una función alterna, llamada fetch, que igualmente descarga información a nuestra carpeta de trabajo, solo que no afecta a las secciones en las que nos encontramos trabajando en ese momento.



Push

1 ↑

El comando push se usa para cargar contenido del repositorio local a un repositorio remoto. El envío es la forma de transferir commits desde tu repositorio local a un repositorio remoto.

Fork

🍴 Fork

La palabra fork se traduce al español, dentro del contexto que nos ocupa, como bifurcación. Cuando hacemos un fork de un repositorio, se hace una copia exacta en crudo (en inglés «bare») del repositorio original que podemos utilizar como un repositorio git cualquiera. Después de hacer fork tendremos dos repositorios git idénticos pero con distinta URL. Justo después de hacer el fork, estos dos repositorios tienen exactamente la misma historia, son una copia idéntica. Finalizado el proceso, tendremos dos repositorios independientes que pueden cada uno evolucionar de forma totalmente autónoma.

Acciones en Github



branches

Ramificaciones del proyecto principal que pueden hacerse para experimentar si hacer daño al repositorio principal.

merge

Acción de fusionar una rama secundaria con el repositorio principal. Esto se puede lograr siempre que no haya conflictos entre archivos y que la persona encargada de administrar el repositorio esté de acuerdo con los cambios realizados.

Acciones en Github



Commit

Commit changes

Cuando se realiza un cambio individual en un solo archivo de nuestro repositorio de Github.



Pull request

Un **pull request** es una petición que el propietario de un fork de un repositorio hace al propietario del repositorio original para que este último incorpore los commits que están en su fork.

Issues

! Issues

Un issue es cuando brota un error en el código de alguien y se lo informa al administrador del repositorio para que este haga los cambios correspondientes.

Acciones en Github



Organizaciones

Las cuentas de *organizaciones* son cuentas compartidas en las cuales negocios o proyectos de código abierto pueden colaborar a través de múltiples proyectos a la vez. Los dueños y los administradores de la organización igualmente, pueden gestionar accesos de los colaboradores a datos y proyectos privados y protegidos por distintos mecanismos de seguridad.

The screenshot shows the GitHub organization page for "México Abierto". The header includes the organization's logo (a green cross icon), name, description ("Código del portal de datos abiertos del Gobierno de la República y otros proyectos"), and contact information. Below the header, there are tabs for "Repositories" (61), "Packages", "People" (17), and "Projects". A search bar, filter buttons for "Type" and "Language", and a "Sort" dropdown are also present. The main content area displays repository cards for "assets", "ckanops-1", and "ceph-ansible". Each card includes the repository name, description, language, and last update date. A sidebar on the right shows "Top languages" (Python, JavaScript, HTML, Ruby, CSS) and a "People" section with a grid of user profiles.

The screenshot shows the GitHub organization page for "Accenture". The header includes the organization's logo, name, description ("Accenture Github site"), and contact information. Below the header, there are tabs for "Repositories" (137), "Packages", "People" (12), and "Projects". A search bar, filter buttons for "Type" and "Language", and a "Sort" dropdown are also present. The main content area displays repository cards for "adop-docker-compose", "reactive-interaction-gateway", and "AmpliGraph". Each card includes the repository name, description, language, and last update date. A sidebar on the right shows "Top languages" (JavaScript, Shell, Groovy, Java, Python) and a "mercury" section with a description ("Mercury - the Post Office for Microservices") and a list of tags: microservices, reactive, kafka, event-stream, spring-boot.

Lecturas



url - Dirección Web	Página
https://github.com/collections/open-journalism	Open Journalism
https://github.com/fivethirtyeight/data	Página de Github de fivethirtyeight
https://fivethirtyeight.com	Página de FiveThirtyEight
https://github.com/eyeseast/awesome-journalism	Awesome Journalism; página de herramientas tecnológicas para periodistas.
https://ijnet.org/es/node/656	How journalists can get started with GitHub
https://github.com/craft2es/githubparaperiodistas	GitHub para periodistas.
https://knightlab.northwestern.edu/2013/06/13/getting-github-why-journalists-should-know-and-use-the-social-coding-site/ https://www.poynter.org/reporting-editing/2015/github-tutorials-and-resources-for-journalists/	Getting GitHub: Why journalists should know and use the social coding site GitHub tutorials
https://learn.r-journalism.com/en/git/github_pages/github-pages/	BEST PRACTICES FOR GITHUB
https://learn.r-journalism.com/en/introduction/	R para periodistas.
https://learn.r-journalism.com/en/git/github/github/	Conectando R con Github
https://greglinch.com/2010/07/quick-thoughts-on-journalism-and-version-control.html	Quick thoughts on journalism and version control (now known as GitHub for News)
https://docs.google.com/presentation/d/1MbltRcOerktc-E26HMDjYj0BO9CTubQWu1Z2bB9CpVY/	Github for journalist (Nicar 2018)
https://medium.com/bbc-visual-and-data-journalism/how-the-bbc-visual-and-data-journalism-team-works-with-graphics-in-r	How the BBC Visual and Data Journalism team works with graphics in R
https://blog.ouseful.info/2017/01/25/data-journalism-units-on-github/?	Data Journalism Units on Github
https://www.machlis.com/nicar19.html	Data Journalism Resources from the NICAR 2019 Conference

Actividad práctica.



Actividades:

1. Abrir nuestra cuenta de Github y ver qué es lo que tiene.
2. Hacer fork a un repositorio interesante.
3. Crear un nuevo repositorio.
4. Clonar el repositorio a mi computadora.
5. Subir archivos a un repositorio.
6. Hacer fetch + comió a mi repositorio remoto.
7. Ver controlador de versiones.
8. Subir una página a Github Pages.
9. Hacer un pull-request a otro repositorio.