



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



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.

GitHub



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 fuentes 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 1... <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> [@datagraphics](https://twitter.com/datagraphics) datagraphics@caltimes.com



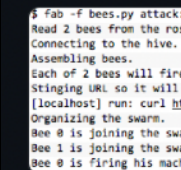
The New York Times

📍 New York, NY <https://nytimes.com> Verified



NPR News Apps team

📍 Washington, DC <http://blog.apps.npr.org> nprapps@npr.org



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](https://twitter.com/gdndevelopers)



FiveThirtyEight

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

Cómo utilizar Github



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



Generar un análisis



Subir código al repositorio



Ponerlo a disposición del público

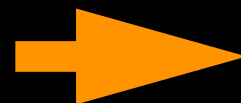
Cómo utilizar Github



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 4 watches, 0 stars, and 0 forks. It contains 9 commits, 1 branch, 0 releases, and 1 contributor. The main branch is 'master'. The repository description is 'Five patients made 8,303 emergency calls in a year' with a link to a BBC news article. The repository contains several files: 'Frequent Callers Data.xlsx', 'README.md', 'average_job_time.csv', 'busiest_month_Oct16_Nov17.csv', 'foirequest.md', and 'frequent_caller_percentage_careplan_in_place.csv'. The README.md file is open, showing the title '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](#).

BuzzFeed NEWS / REPORTING TO YOU | BuzzFeed | Videos | Quizzes | Tasty | As/Is | More ▾

BuzzFeed News Trained A Computer To Search For Hidden Spy Planes. This Is What We Found.

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.

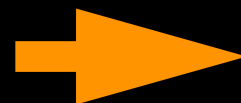
Originally posted on August 7, 2017, at 12:33 p.m.
Updated on August 8, 2017, at 9:47 a.m.

Peter Aldhous
BuzzFeed News Reporter

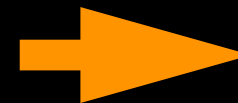
A secret spy plane operated by the US Marshals hunted drug cartel kingpins in Mexico. A military contractor that tracks terrorists in Africa is also flying surveillance aircraft over US cities. In two stories published last week, BuzzFeed News revealed the activities of aircraft that their operators didn't want to discuss.

These discoveries came not from tip-offs from anonymous sources, but by training a computer to recognize known spy planes, then setting it loose on large quantities of flight-tracking data

Generar un análisis



Subir código al repositorio



Ponerlo a disposición del público

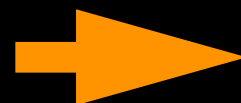
Cómo utilizar Github



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

The screenshot shows a GitHub repository page for 'BuzzFeedNews / 2017-08-spy-plane-finder'. The repository has 7 watchers, 47 stars, and 16 forks. It contains 11 commits, 2 branches, 0 releases, and 1 contributor. The repository description states: 'The data and analysis referenced in the Aug. 7, 2017 BuzzFeed News article, "BuzzFeed News Trained A Computer To Search For Hidden Spy Planes. This Is What We Found." <https://www.buzzfeed.com/peteraldhous/hidden-spy-planes>'. The repository contains a 'data' directory, 'README.md', 'index.Rmd', and 'index.html'. The 'README.md' file is selected and shows the title 'BuzzFeed News Trained A Computer To Search For Hidden Spy Planes. This Is What We Found.' and a description: '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.













Conceptos clave



🕒 29 commits

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		Visualización	Grupo	Compartir	Agregar etiquetas	Acción	Buscar
Nombre		Fecha de modificación	Tamaño	Clase			
	Copia de tesis	hoy 19:03	12 KB	Micros...(docx)			
	Copia de tesis 2	hoy 19:03	12 KB	Micros...(docx)			
	tesis	hoy 19:03	12 KB	Micros...(docx)			
	Tesis Versión Definitiva	hoy 19:03	12 KB	Micros...(docx)			
	Tesis Versión Final	hoy 19:03	12 KB	Micros...(docx)			
	Tesis Versión Final Final	hoy 19:03	12 KB	Micros...(docx)			
	Tesis Versión Final Final_2	hoy 19:03	12 KB	Micros...(docx)			
	Tesis_v2	hoy 19:03	12 KB	Micros...(docx)			
	Tesis_v3	hoy 19:03	12 KB	Micros...(docx)			
	Tesis_v4	hoy 19:03	12 KB	Micros...(docx)			

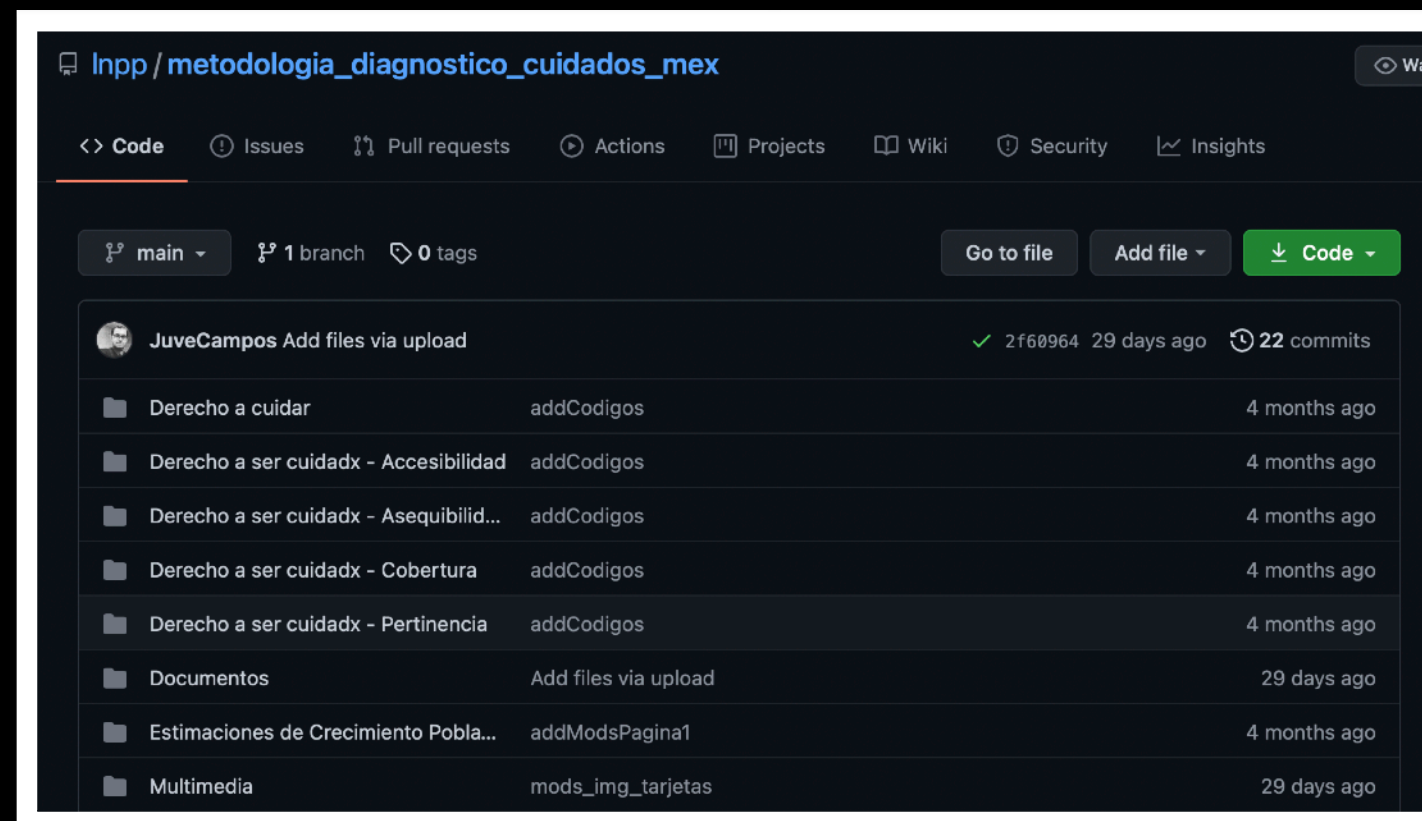
Conceptos clave



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.

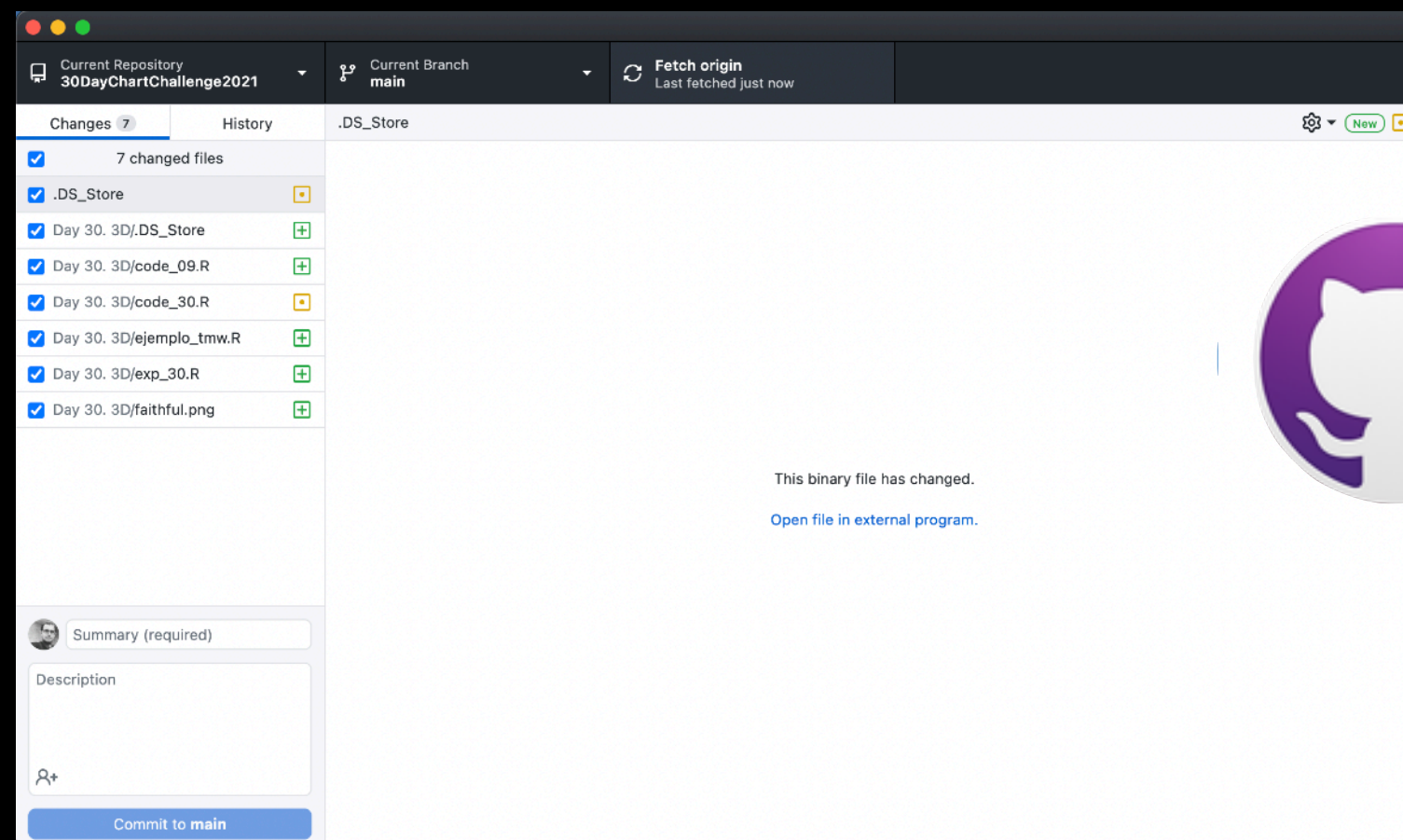


Conceptos clave



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

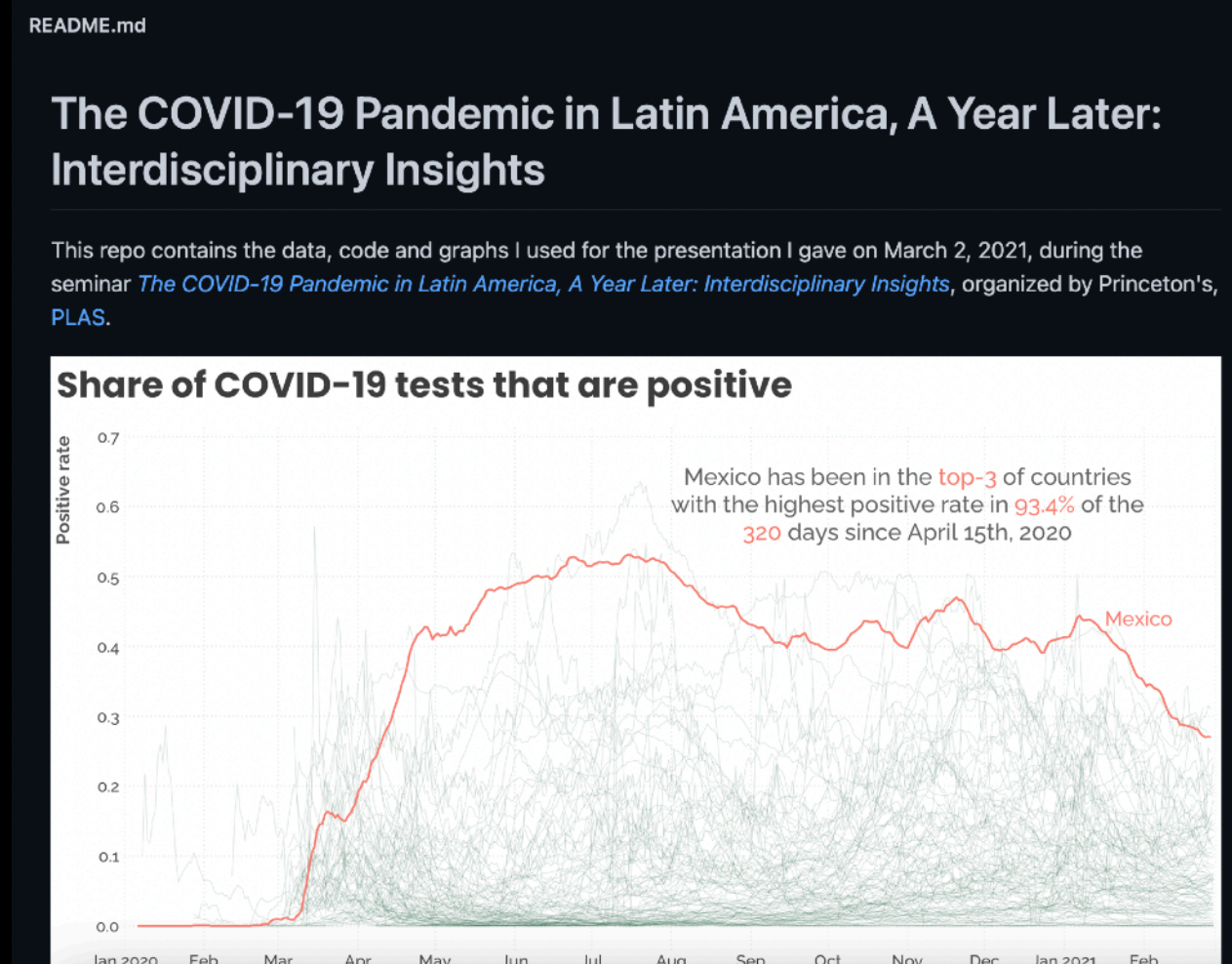


Proyecto de metodología, diagnóstico y propuesta de cuidados a nivel municipal

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](#)".



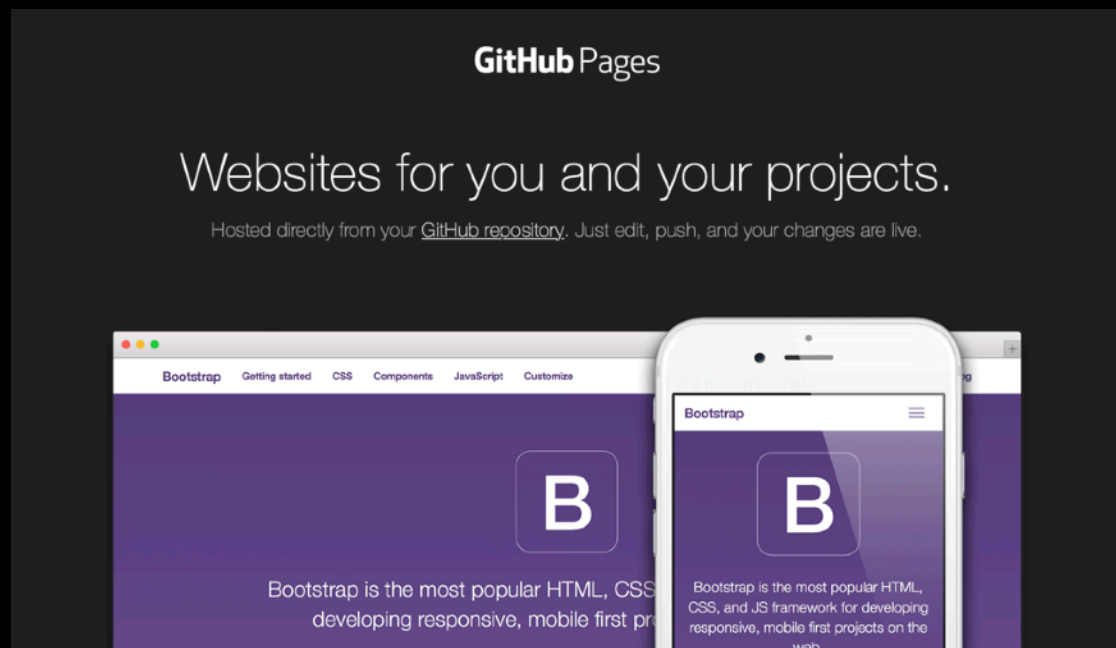
Conceptos clave



github-pages

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.



2021

#30DayChartChallenge 2

Presentaciones

Juvenal Campos
2/27/2020

Sitio de almacenamiento de Páginas Web

Clase de Periodismo de Datos 2021 - MPPP

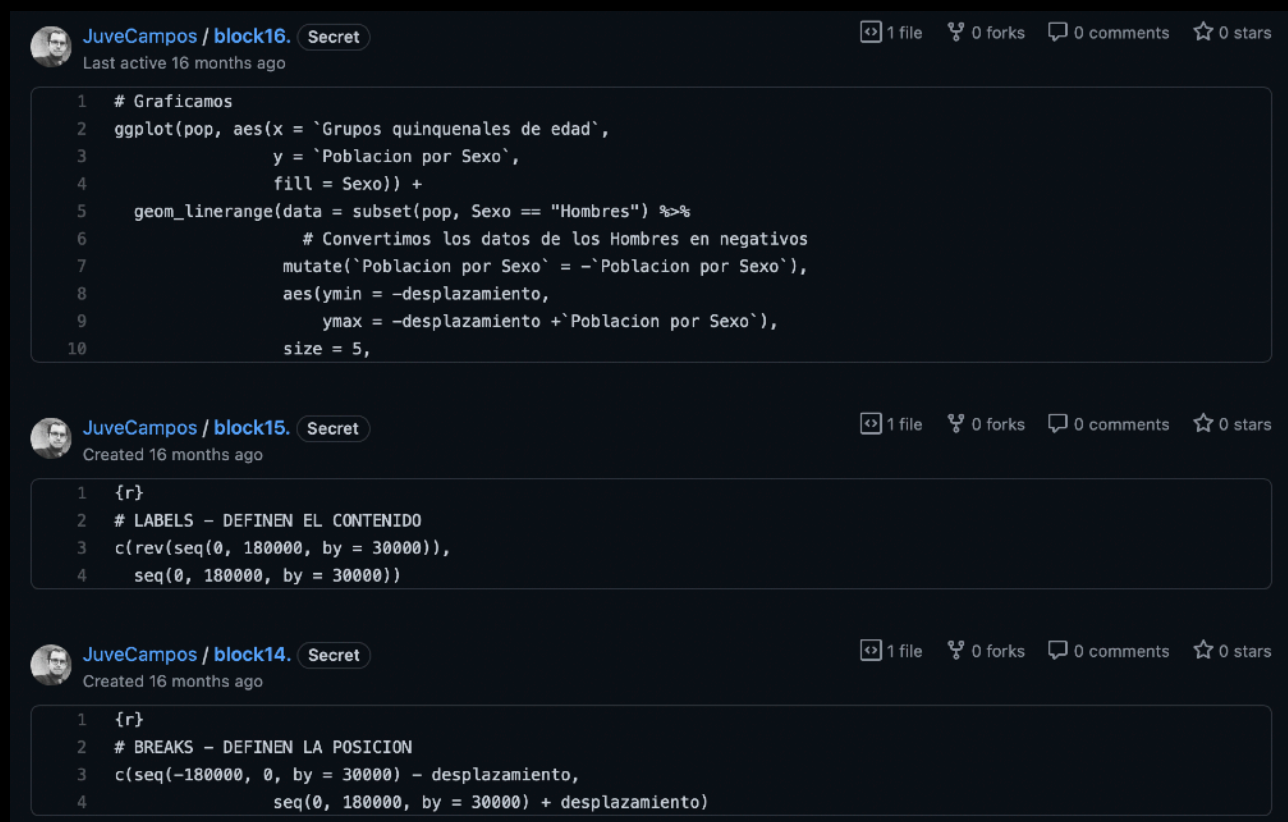
- Sesión 1. Introducción a RMarkdown
- Sesión 3. Expresiones regulares con rebus
- Sesión 4. Mapas en ggplot2
- Sesión 5. Mapas interactivos con leaflet
- Sesión 6. Visualización interactiva en R
- Sesión 7. Datos Abiertos y Descargas Automatizadas
- Sesión 8 y 9. Descarga de datos de APIs
- Sesión 10.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



The screenshot displays three GitHub Gists created by the user JuveCampos. Each gist is labeled as 'Secret' and has 1 file, 0 forks, 0 comments, and 0 stars. The first gist, 'block16', is a ggplot2 script for a geom_linerange plot. The second, 'block15', contains R code for creating a sequence of values. The third, 'block14', also contains R code for creating a sequence of values with specific breaks.

```
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,
```

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

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

Gists de JuveCampos.



The screenshot displays two GitHub Gists created by the user jjsantos01. The first gist, 'dgis_conexion.ipynb', is a Jupyter Notebook titled 'Pasos para conectarse a los cubos dinámicos de la DGIS de la Secretaría de Salud'. It includes instructions on installing 'adodbapi' and connecting to a database. The second gist, 'parejas.ipynb', is a Jupyter Notebook titled 'Calcula hogares con parejas homosexuales a partir del censo 2020'. It includes code for importing libraries and a link to a news article.

```
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
```

```
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
```

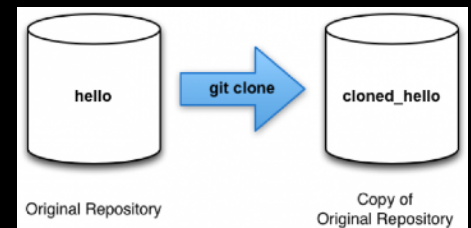
Gists de jjsantos01.



Colaboración en Github.

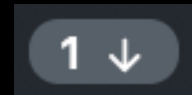
Para colaborar en github, hay que saber algunos de los términos necesarios para entender que acciones estamos llevando a cabo.

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



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 encontremos trabajando en ese momento.

Acciones en Github



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



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.

México Abierto
Código del portal de datos abiertos del Gobierno de la República y otros proyectos
📍 México 🌐 <http://datos.gob.mx/> ✉ escuadron@datos.gob.mx

Repositories 61 Packages People 17 Projects

Find a repository... Type Language Sort

assets
Un repositorio de recursos comunes para nuestros proyectos
👤 0 ⭐ 0 📄 5 📁 0 Updated on 15 Feb

ckanops-1
Forked from urkonn/ckanops
ALL YOUR DATASETS ARE BELONG TO US
🔗 Python 📄 MIT 👤 5 ⭐ 0 📄 0 📁 0 Updated on 25 Dec 2020

ceph-ansible
Forked from ceph/ceph-ansible
Ansible playbooks to deploy Ceph, the distributed filesystem.

Top languages
Python JavaScript HTML Ruby CSS

People 17 >

Accenture
Accenture Github site
🌐 <https://accenture.github.com>

Repositories 137 Packages People 12 Projects

Find a repository... Type Language Sort

Pinned repositories

- adop-docker-compose** (Archived)
Talk to us on Gitter: <https://gitter.im/Accenture/ADOP>
🟢 Shell ⭐ 764 👤 589
- reactive-interaction-gateway**
Create low-latency, interactive user experiences for stateless microservices.
🟡 Elixir ⭐ 473 👤 56
- AmpliGraph**
Python library for Representation Learning on Knowledge Graphs <https://docs.ampligraph.org>
🟢 Python ⭐ 1.5k 👤 166

mercury
Mercury - the Post Office for Microservices
microservices reactive kafka event-stream spring-boot

Top languages
JavaScript Shell Groovy Java Python

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/	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/	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/1MbItRcOerKtc-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.