Data Analysis Toolkit

Sebastián Ocampo Palacios
Anna Karina Pérez Peña

Introduction

In this presentation we will present some useful tools for creating and sharing data visualizations. Mainly, we will explore:

R Graph Gallery

stackoverflow

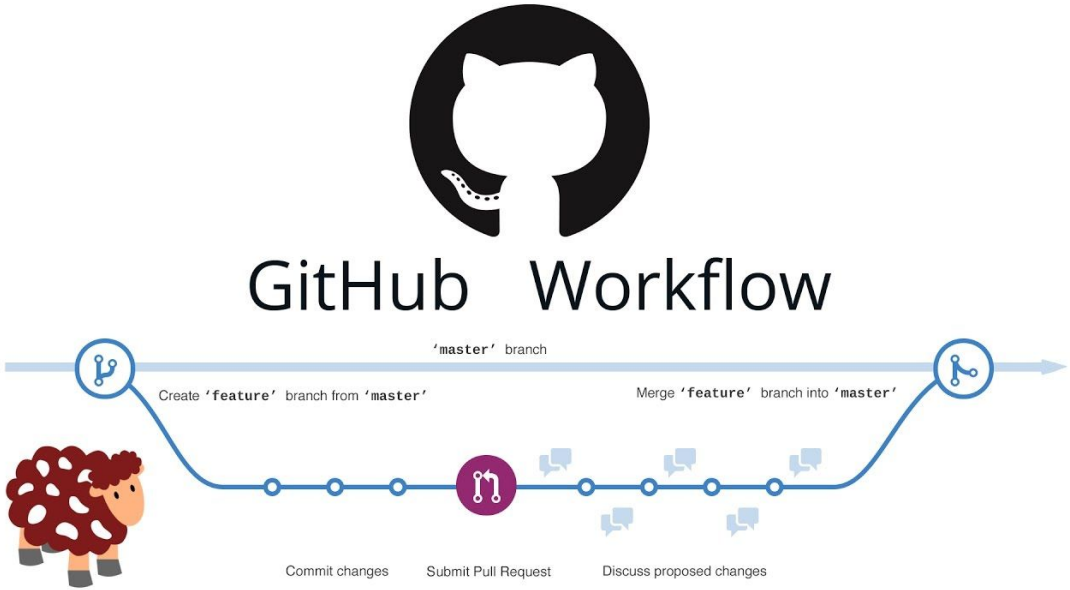
Github

Github provides an interface to collaborate on code programming.

It also allows sharing of datasets, codes and libraries.

Main Features

- Git
- Branches
- Commits
- Pull Requests



Concepts

Git is the software.

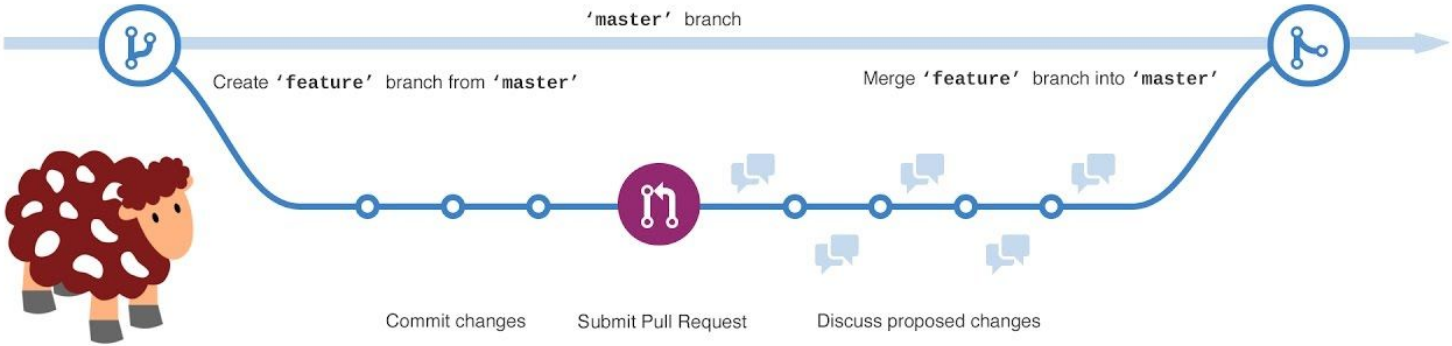
Branches allow parallel versions of a same file.

Pull Requests display the changes done to a branch for their discussion and approval.

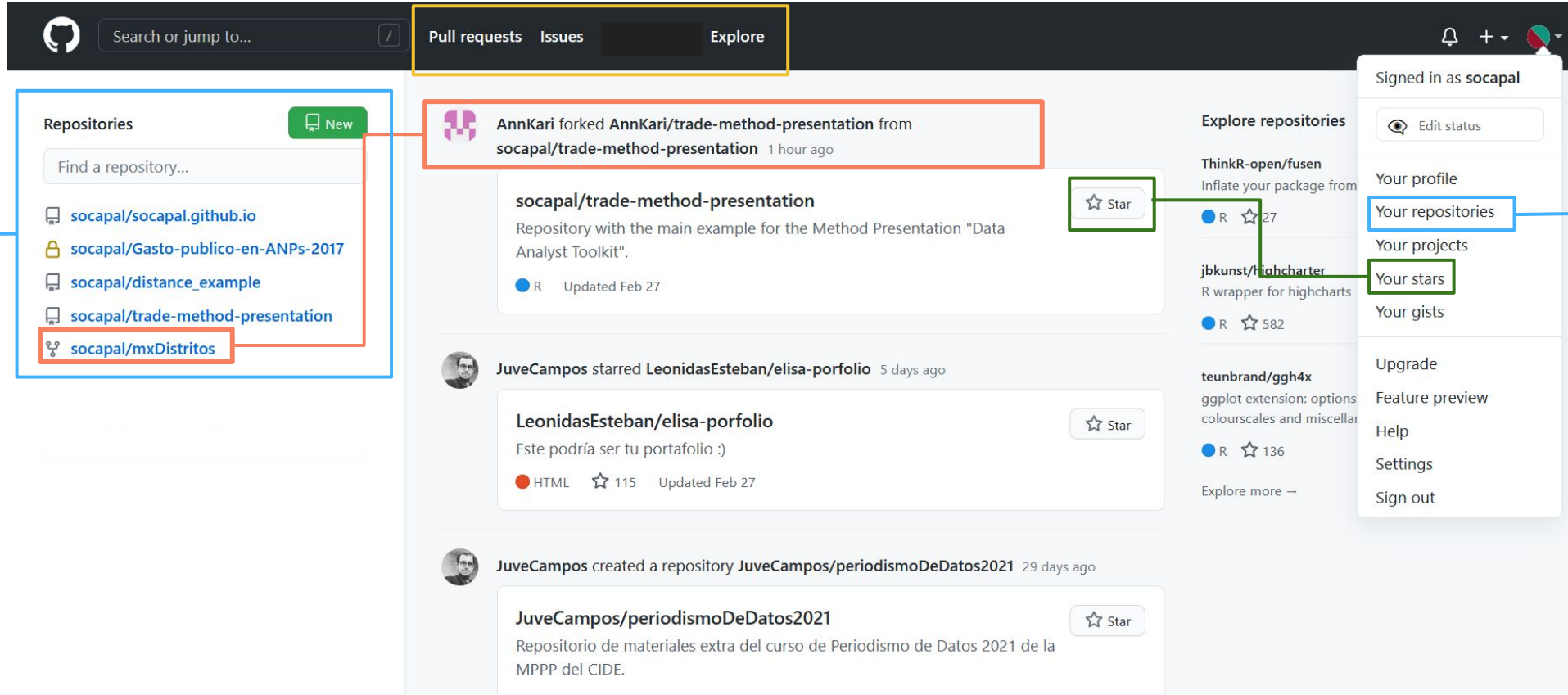
Commits serve as snapshots of the repository at specific times.

Repositories contain project files and each file's version history.

GitHub Workflow



Interface


Search or jump to...
Pull requests
Issues
Explore
Repositories
New
Find a repository...
socapal/socapal.github.io
socapal/Gasto-publico-en-ANPs-2017
socapal/distance_example
socapal/trade-method-presentation
socapal/mxDistritos
AnnKari forked AnnKari/trade-method-presentation from socapal/trade-method-presentation 1 hour ago
socapal/trade-method-presentation
Repository with the main example for the Method Presentation "Data Analyst Toolkit".
R Updated Feb 27
Star
JuveCampos starred LeonidasEsteban/elisa-porfolio 5 days ago
LeonidasEsteban/elisa-porfolio
Este podría ser tu portafolio :)
HTML 115 Updated Feb 27
Star
JuveCampos created a repository JuveCampos/periodismoDeDatos2021 29 days ago
JuveCampos/periodismoDeDatos2021
Repositorio de materiales extra del curso de Periodismo de Datos 2021 de la MPPP del CIDE.
Star
Explore repositories
ThinkR-open/fusen
Inflate your package from
R 27
jbkunst/highcharter
R wrapper for highcharts
R 582
teunbrand/ggh4x
ggplot extension: options
colourscales and miscella
R 136
Explore more →
Signed in as socapal
Edit status
Your profile
Your repositories
Your projects
Your stars
Your gists
Upgrade
Feature preview
Help
Settings
Sign out

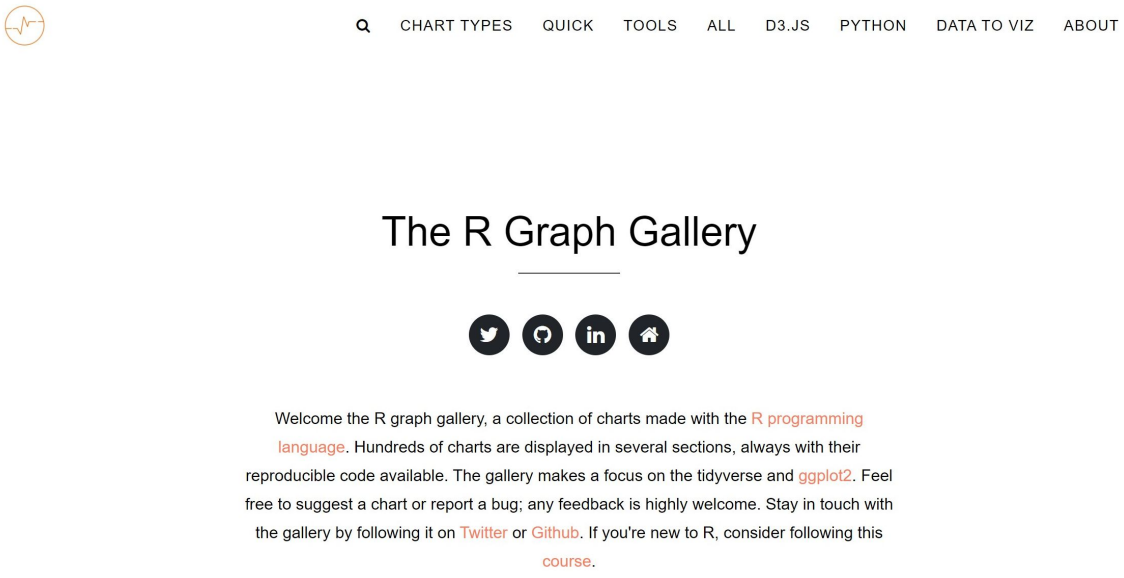
A Guided Example

Creating a Data Visualization

First, we are going to start looking for examples of data visualizations that can be useful as a basis for our work.

R Graph Gallery

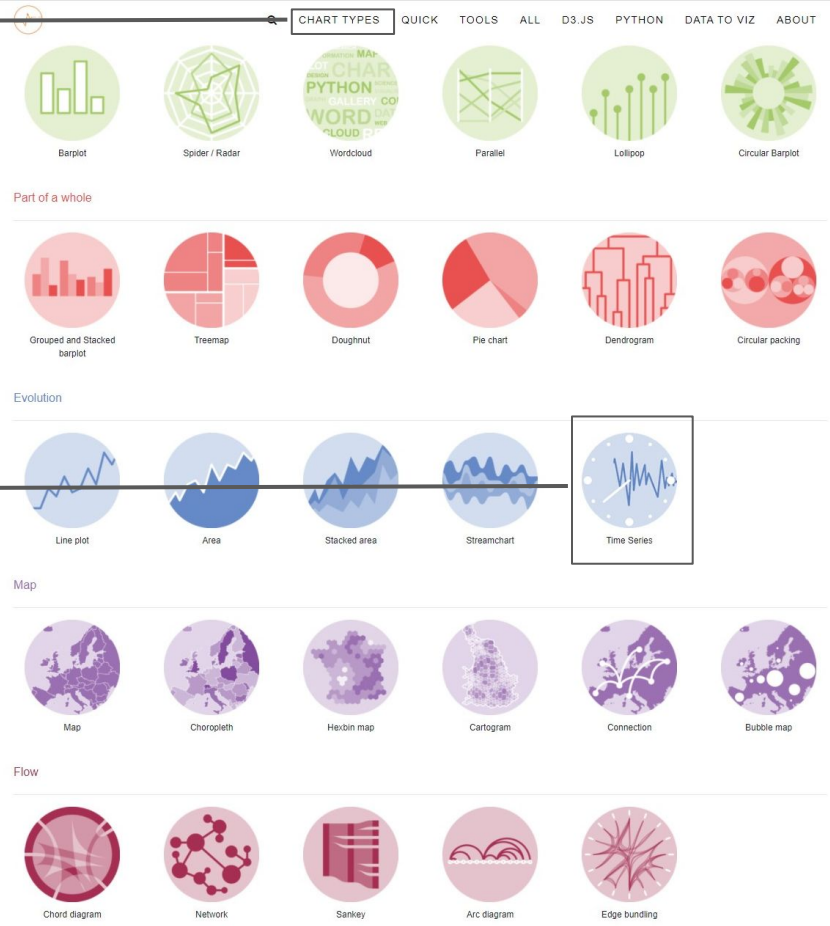
Home screen



If we want to make a data visualization in R, Python or Java, R Graph Gallery will help us to find some ideas and also the complete command of a specific graph.

Chart types

Shows the different types of charts in the collection

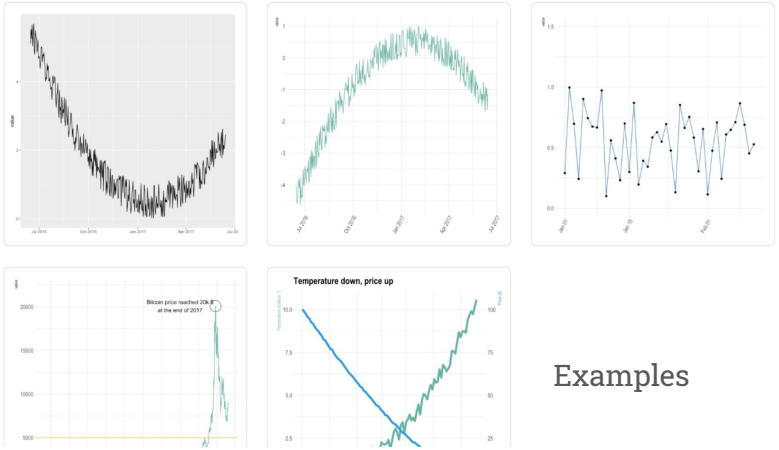


Times Series

TIME SERIES WITH GGPLOT2

Brief description of the packages and commands used

`ggplot2` offers great features when it comes to visualize time series. The `date` format will be recognized automatically, resulting in neat X axis labels. The `scale_x_data()` makes it a breeze to customize those labels. Last but not least, `plotly` can turn the resulting chart interactive in one more line of code.



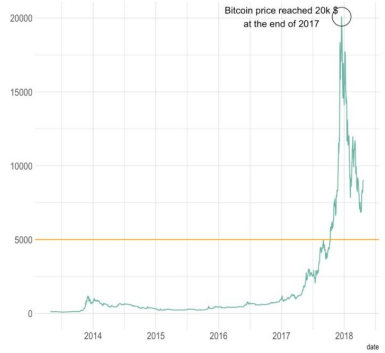
Examples

Example

The `ggplot2` package recognizes the `date` format and automatically uses a specific type of X axis. If the time variable isn't at the `date` format, this won't work. Always check with `str(data)` how variables are understood by R. If not read as a date, use lubridate to convert it. Read more about this here.

On the chart beside, dates are displayed using a neat format: month + year.

Note: the gallery offers a section dedicated to line charts.



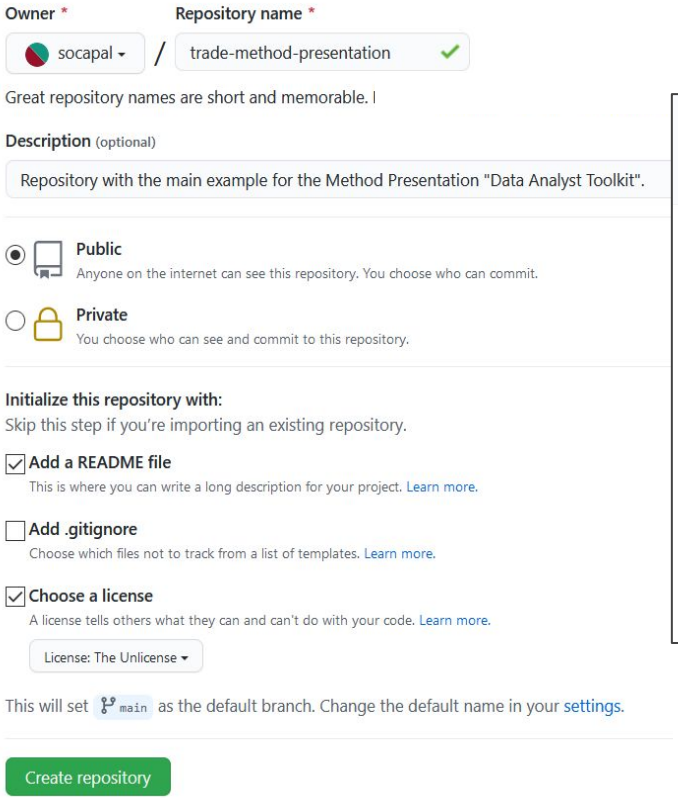

```r
# libraries
library(ggplot2)
library(dplyr)
library(plotly)
library(hrbrthemes)

# Load dataset from github
data <- read.table("https://raw.githubusercontent.com/holtzy/data_to_viz/master/Example_dataset/3_TwoNumOrdered.csv", header=T)
data$date <- as.Date(data$date)

# plot
data %>%
  ggplot( aes(x=date, y=value)) +
    geom_line(color="#69b3a2") +
    ylim(0,22000) +
    annotate(geom="text", x=as.Date("2017-01-01"), y=20089,
             label="Bitcoin price reached 20k $\nat the end of 2017") +
    annotate(geom="point", x=as.Date("2017-12-17"), y=20089, size=10, shape=21, fill="transparent") +
    geom_hline(yintercept=5000, color="orange", size=.5) +
    theme_ipsum()
```

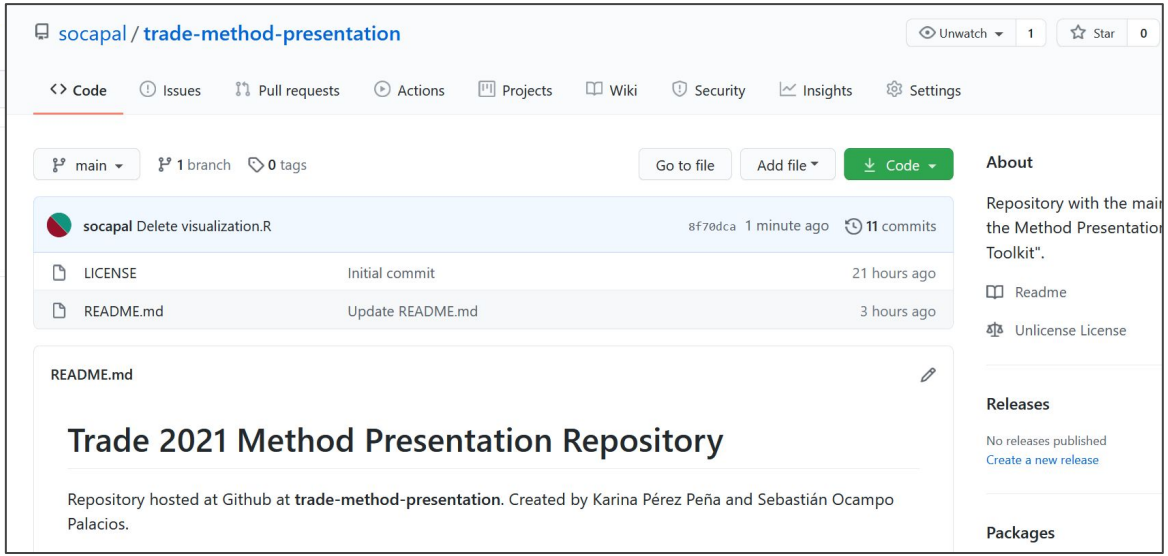
Specific functionalities of the package and the commands used

Code in R

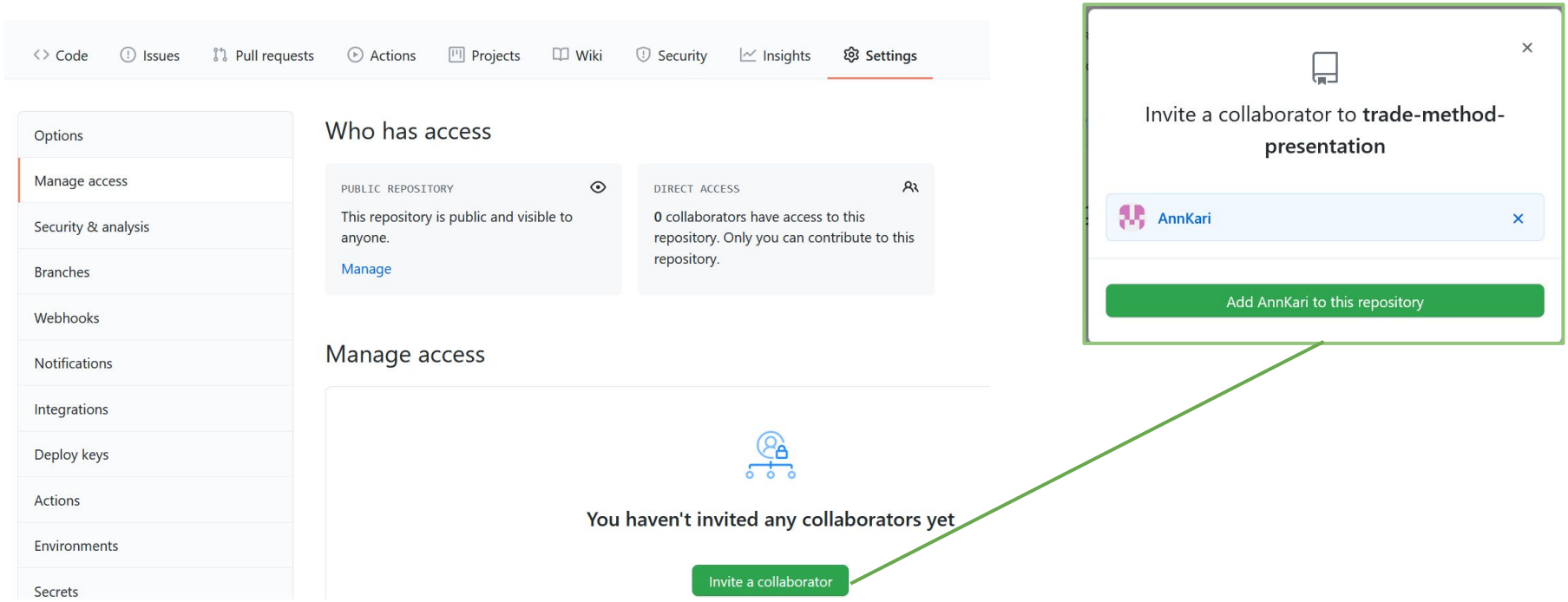
We then proceed to set the Github repository and workplace.

Step 2. Creating a Github Repository





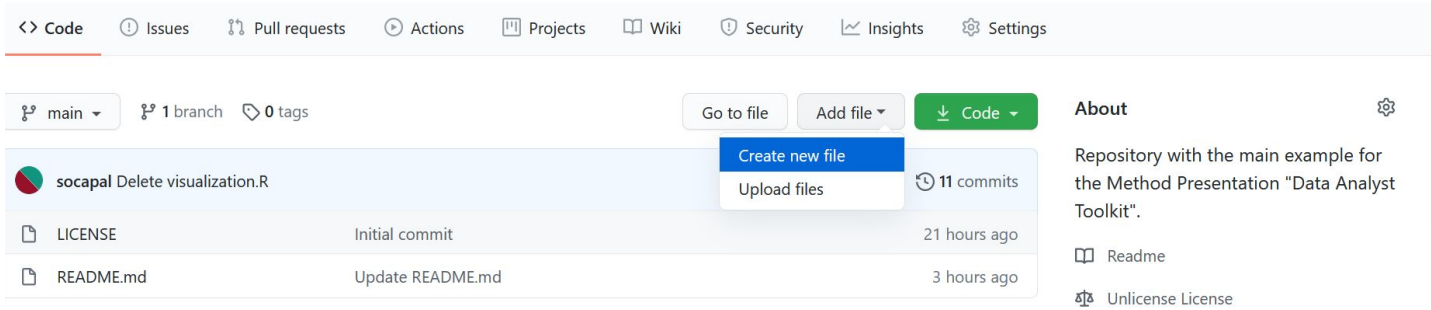
Step 3. Adding a Collaborator

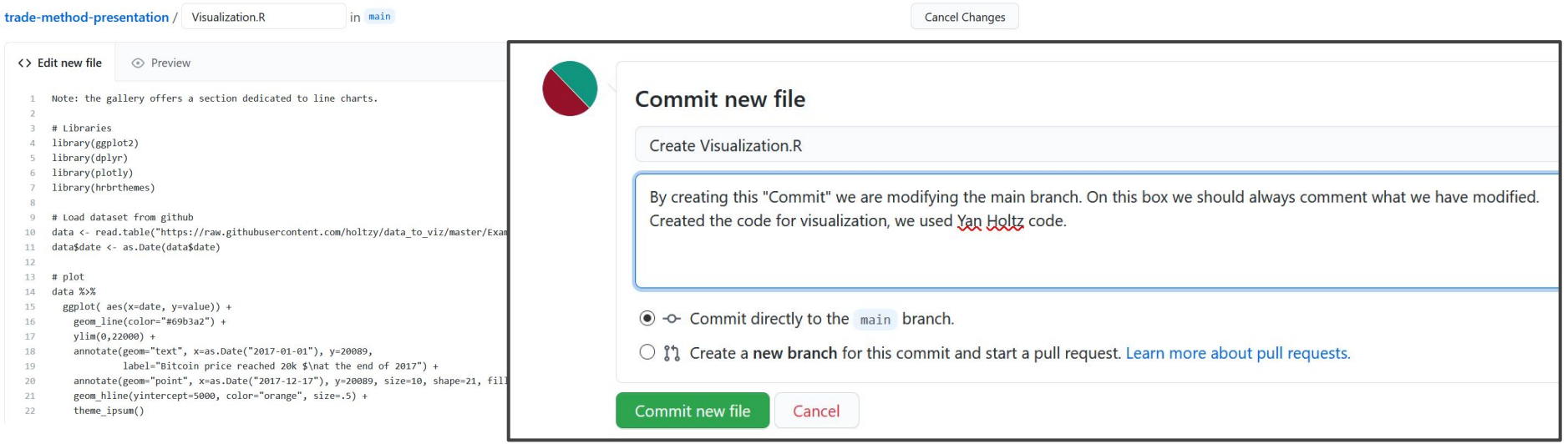


Invitations can be accepted through the user's mail.

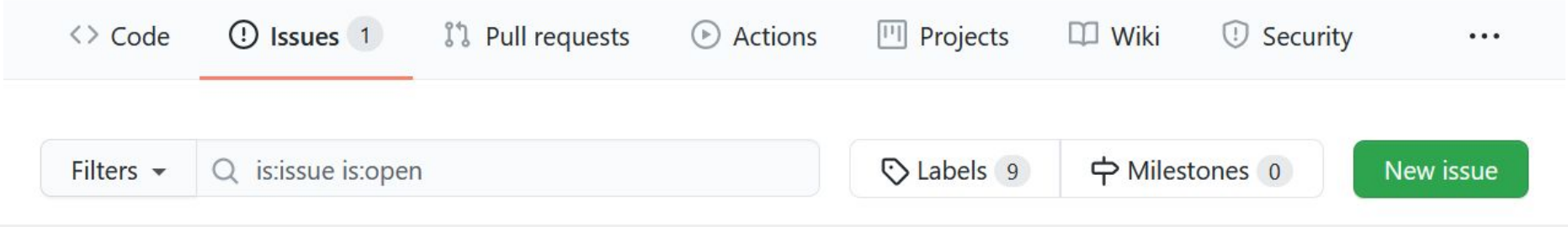
We are now ready to code collaboratively.

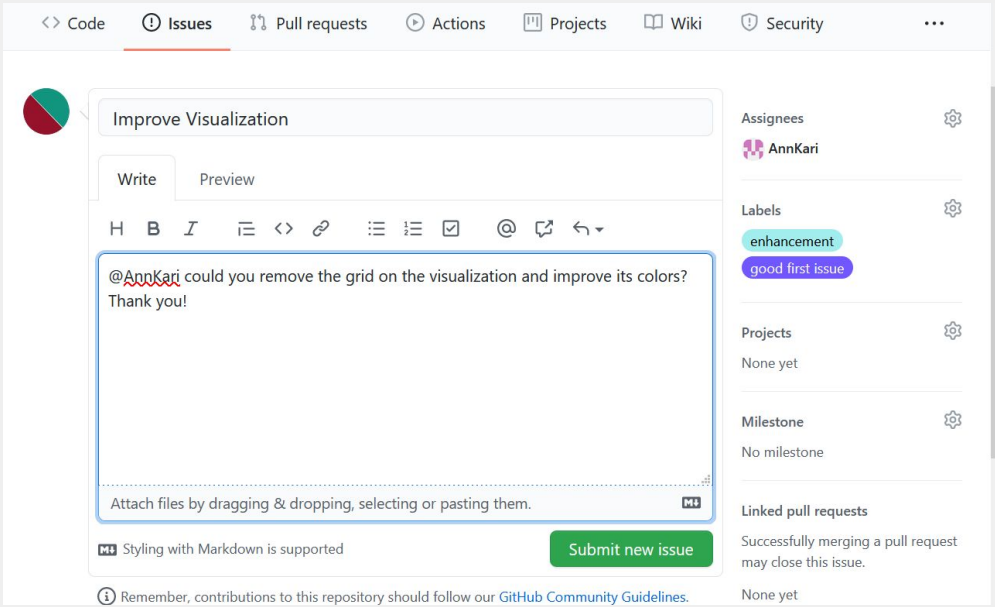
Step 4. Creating the code file

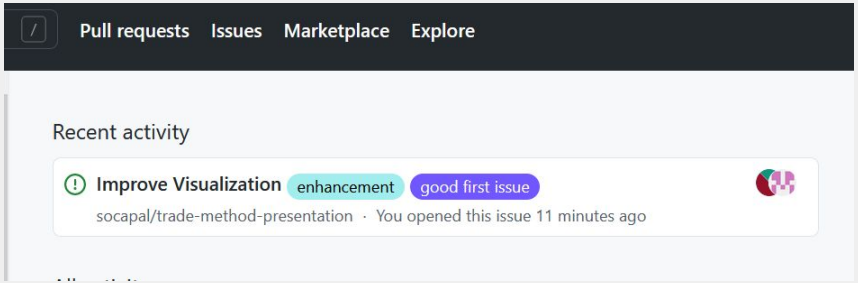




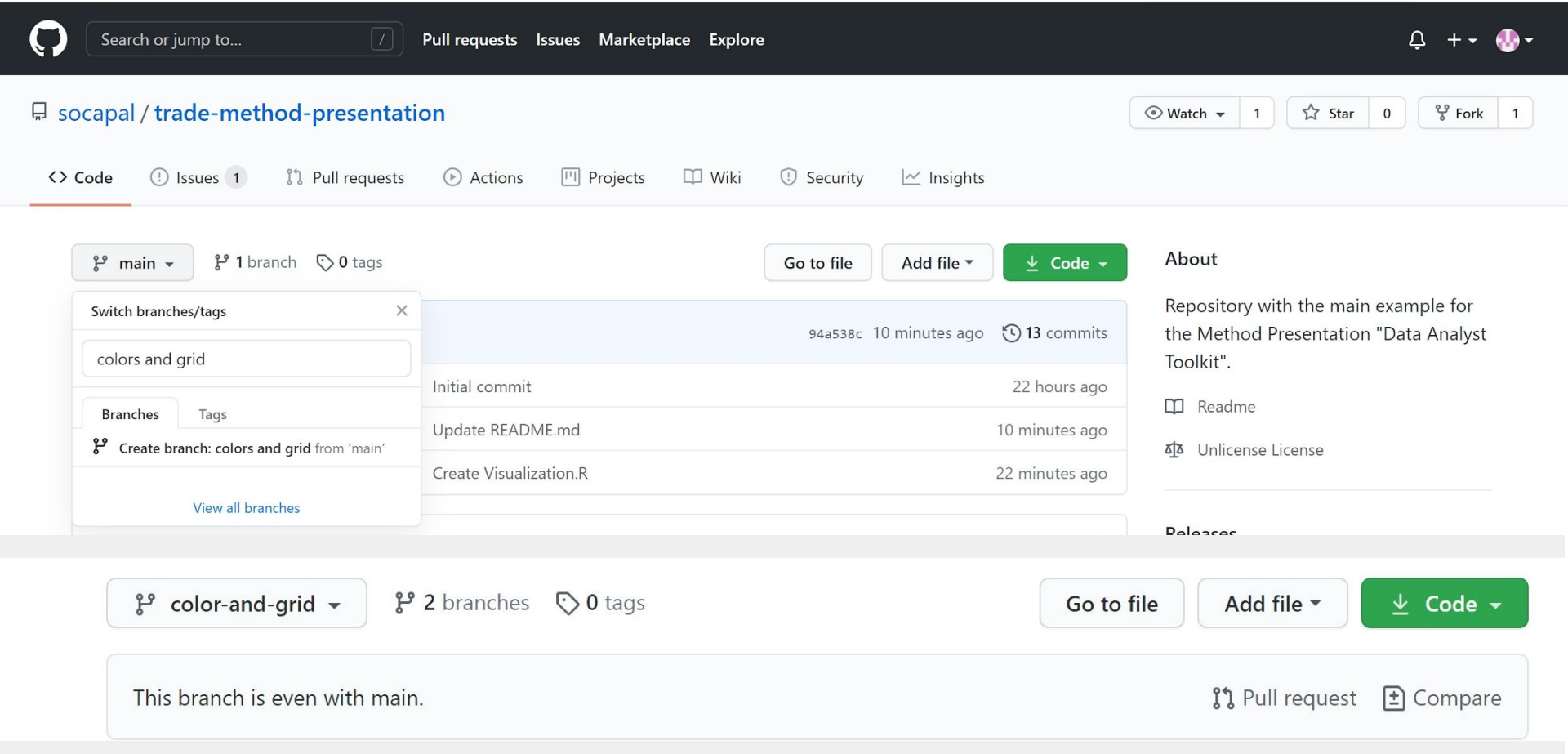
Step 5. Creating an issue to your collaborator



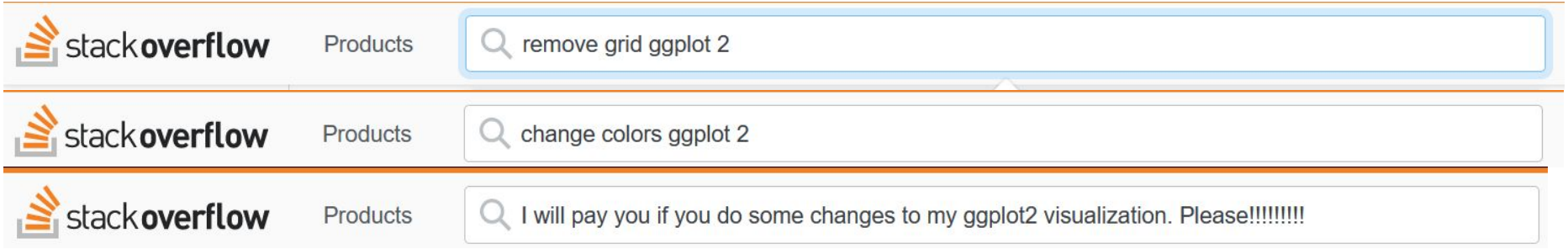


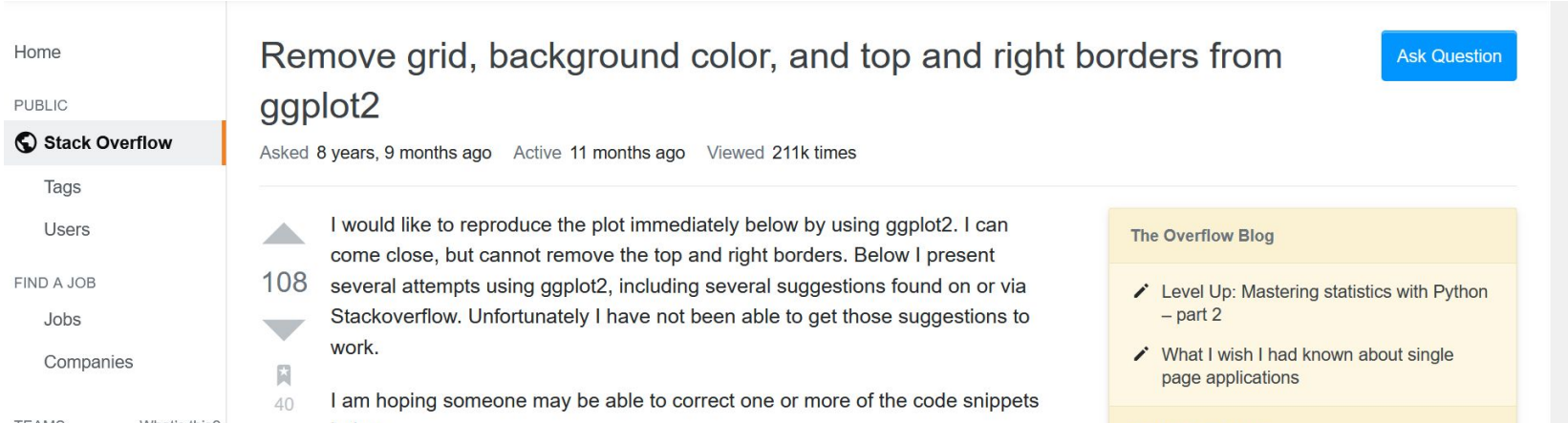


Step 6. Implementing changes through a new branch

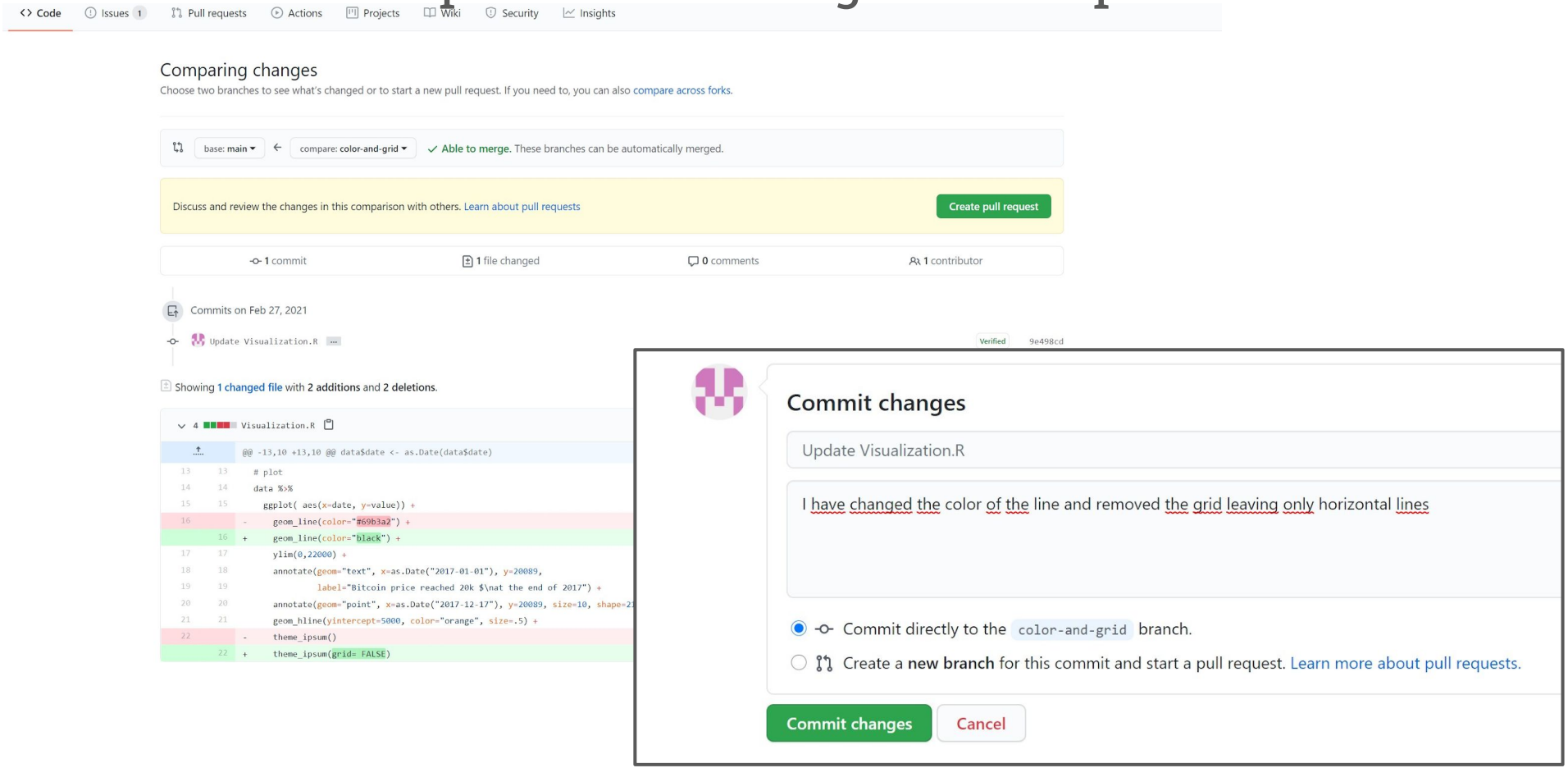


Step 7. Search questions, find answers.

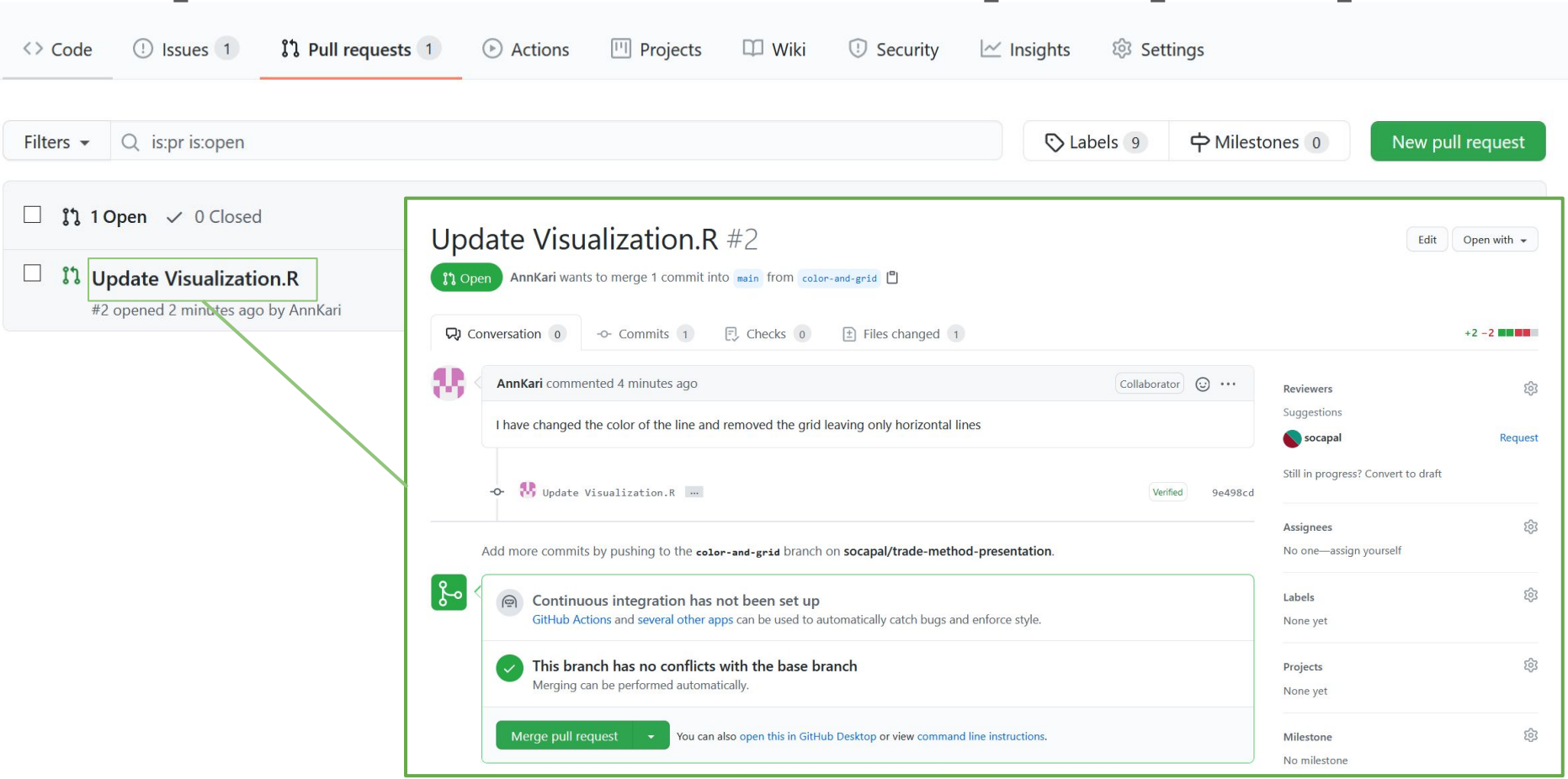




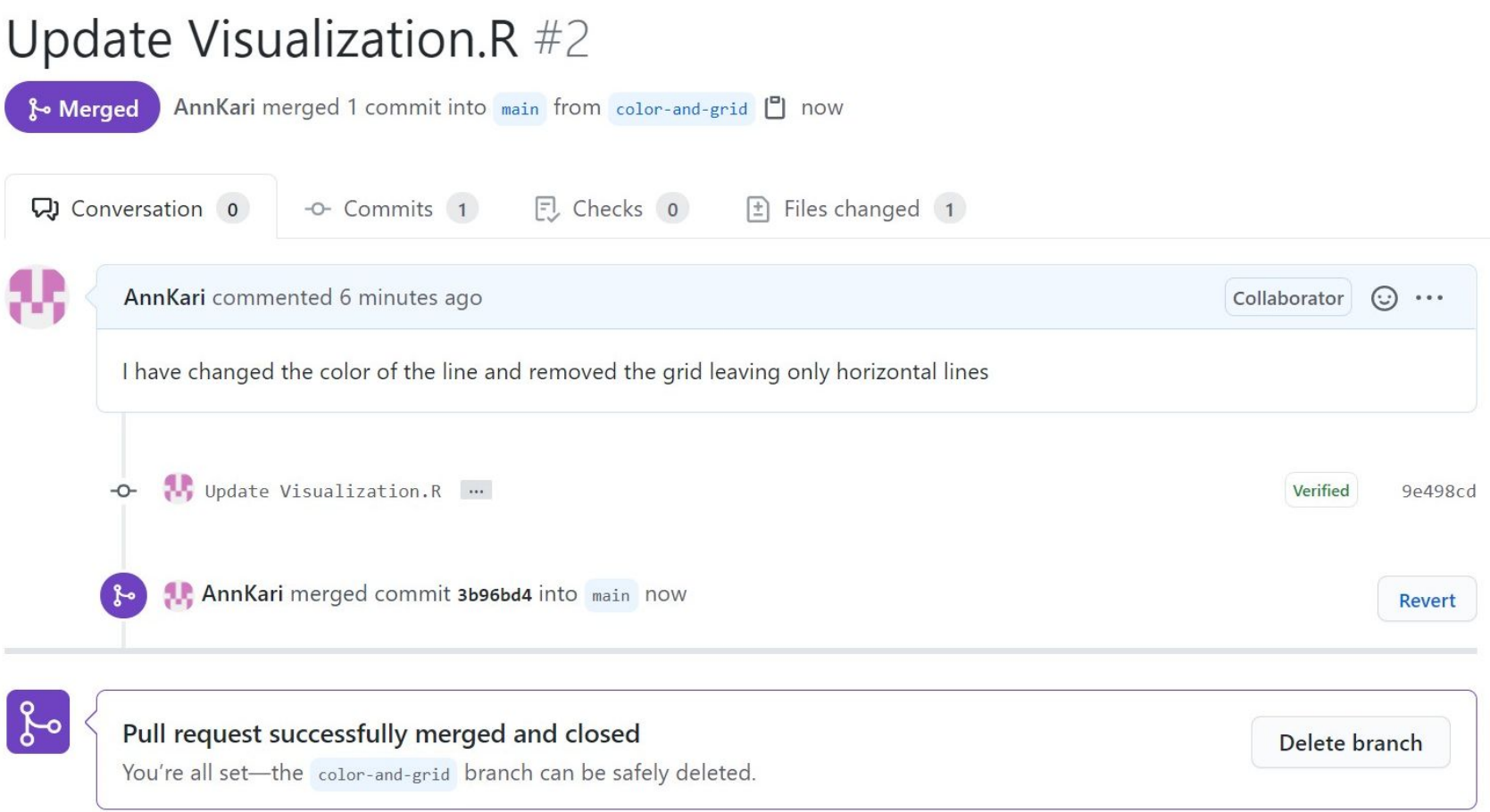
Step 8. Make changes & compare



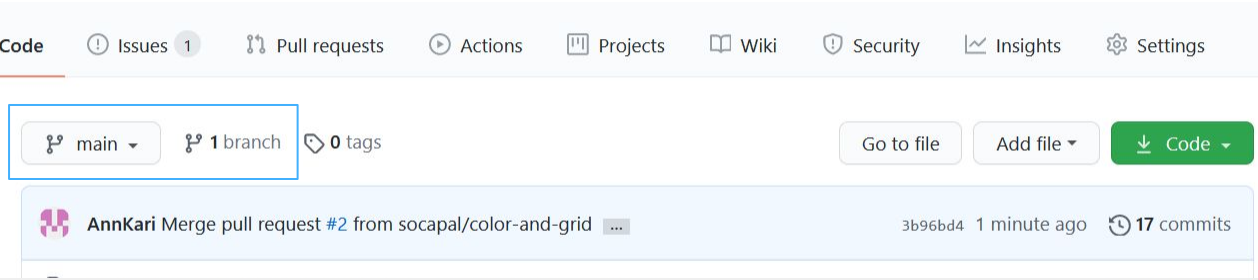
Step 9. Collaborator can now accept the pull request

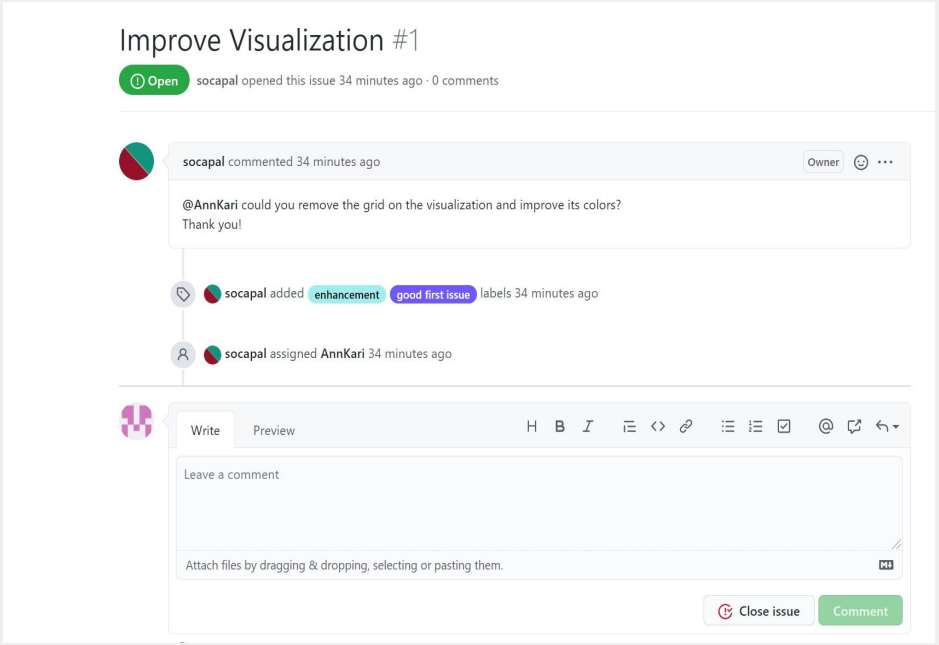


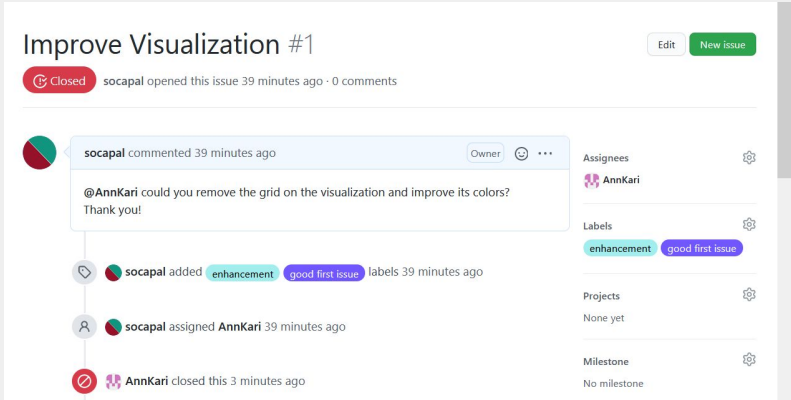
Step 10. Merge branches, delete the redundant branch



Step 11. Close Issue.







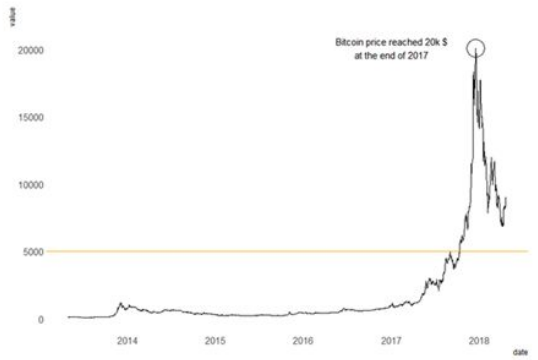
Finally, let’s publish our graph!

Step 12. Publication

r/dataisbeautiful · Posted by [redacted] 3 minutes ago

[OC] Bitcoin price from 2014 to mid 2018

OC



3 Comments Share Save Hide ··· 75% Upvoted

[redacted] 1 minute ago

Hello! This is a sample post. All posts must comply both to Reddit and r/dataisbeautiful rules, _please make sure you read them_. Plagarism is not tolerated.

r/dataisbeautiful is a place to share and discuss visual representations of data: Graphs, charts, maps, etc. Posting allows other people to comment and make suggestions on your graphs. As one of r/dataisbeautiful rules states, all visualizations must provide both the source of the data and the tool/software used:

Data Source. Yan Holtz. Holtz, Yan. s. f. «Line Chart Annotation with Ggplot2». Available at the example data set 3 on https://github.com/holtzy/data_to_viz.
Tools used. RStudio, ggplot2 (Code provided by Yan Holtz).

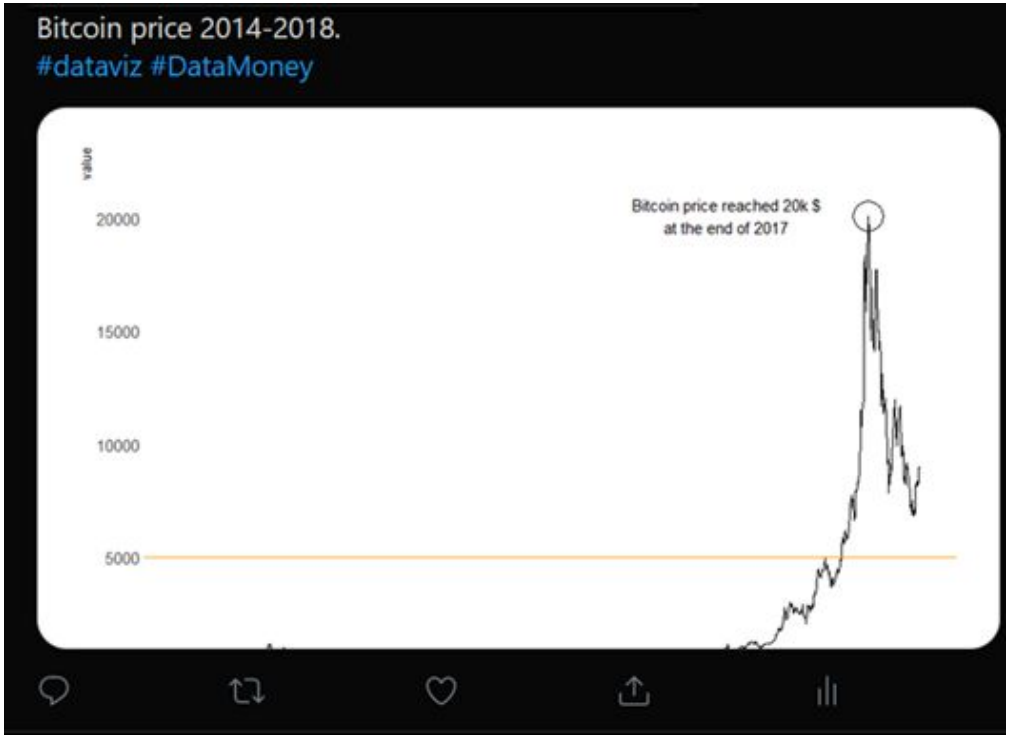
1 Reply Share Save Edit ···

socapal 1 point · 1 minute ago · 0 children

[redacted] just now

This is a simple graph, but it does looks nice! You did not provide information on which currency you used (I assume it to be US dollars) neither you specified on what base year you deflacted that currency!

Vote Reply Give Award Share Report Save



Consulted sources

“Creating a Repository on GitHub - GitHub Docs.” Accessed February 27, 2021. https://docs.github.com/en/github/creating-cloning-and-archiving-repositories/creating-a-repository-on-github.

Holtz, Yan. «Line Chart Annotation with Ggplot2». Accedido 27 de febrero de 2021. https://www.r-graph-gallery.com/line_chart_annotation.html.

“Introduction to GitHub.” *GitHub Learning Lab*. Accessed February 27, 2021. https://lab.github.com/githubtraining/introduction-to-github.

“R - Remove Grid, Background Color, and Top and Right Borders from Ggplot2.” *Stack Overflow*. Accessed February 27, 2021. https://stackoverflow.com/questions/10861773/remove-grid-background-color-and-top-and-right-borders-from-ggplot2.

All Steps

1. Search for a sample or inspiration
2. Create a Github repository
3. Add a collaborator
4. Create a file
5. Create an issue for your collaborator
6. Implement changes through a new branch
7. Search for questions, find answers
8. Make changes & compare
9. Ask your collaborator to accept pull request
10. Merge branches, delete redundant ones
11. Close issue
12. Publish!