

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



data·visualization
at Twitter

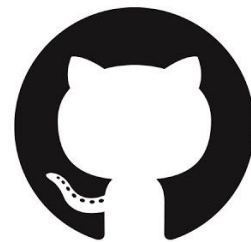
Github

Github provides an interface to collaborate on code programming.

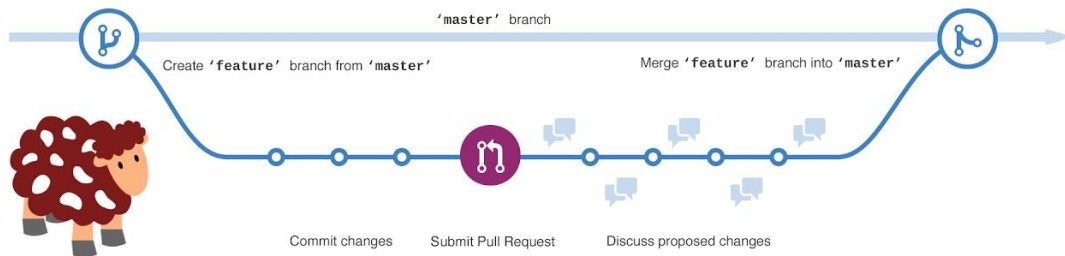
It also allows sharing of datasets, codes and libraries.

Main Features

- Git
- Branches
- Commits
- Pull Requests



GitHub Workflow



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.

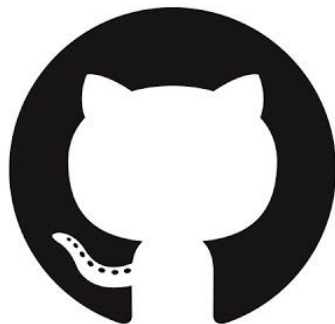
Commits serve as snapshots of the repository at specific times.

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

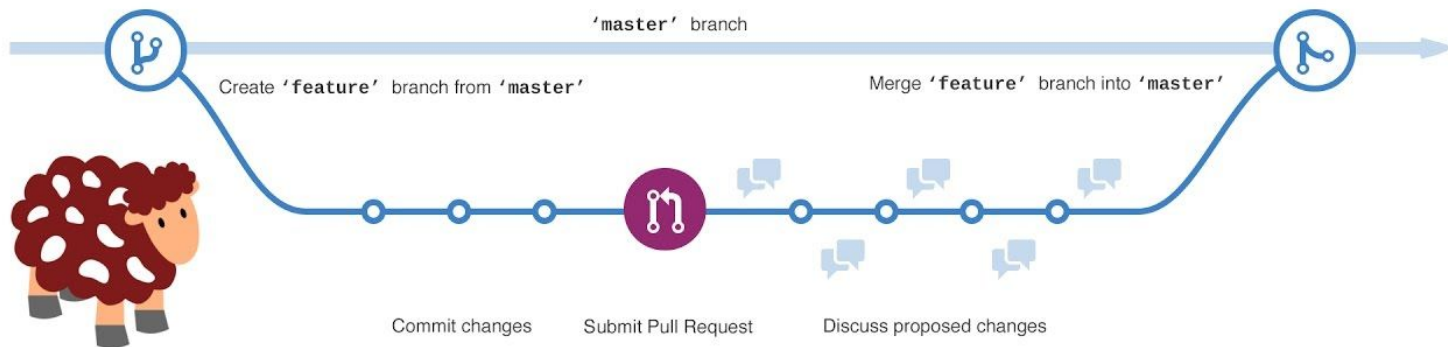
Main Features

● Git ● Branches

● Commits ● Pull Requests



GitHub Workflow



Interface

The screenshot shows the GitHub web interface with several annotations:

- Header:** The GitHub logo, a search bar with the text "Search or jump to...", and navigation links for "Pull requests", "Issues", and "Explore".
- Left Sidebar:** A "Repositories" section with a "New" button and a list of repositories: "socapal/socapal.github.io", "socapal/Gasto-publico-en-ANPs-2017", "socapal/distance_example", "socapal/trade-method-presentation", and "socapal/mxDistritos". The last repository is highlighted with a red box.
- Main Content Area:** A list of repositories. The first entry, "AnnKari forked AnnKari/trade-method-presentation from socapal/trade-method-presentation", is highlighted with a red box. Below it, the repository "socapal/trade-method-presentation" is shown with a "Star" button highlighted by a green box. Further down, "JuveCampos starred LeonidasEsteban/elisa-portfolio" and "JuveCampos created a repository JuveCampos/periodismoDeDatos2021" are listed.
- Right Sidebar:** An "Explore repositories" section showing "ThinkR-open/fusen" and "jbkunst/highcharter".
- User Menu:** A dropdown menu for the user "socapal" is open, showing options like "Your profile", "Your repositories" (highlighted with a blue box), "Your projects", "Your stars" (highlighted with a green box), "Your gists", "Upgrade", "Feature preview", "Help", "Settings", and "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



Q CHART TYPES QUICK TOOLS ALL D3.JS PYTHON DATA TO VIZ ABOUT

The R Graph Gallery



Welcome the R graph gallery, a collection of charts made with the [R programming language](#). Hundreds of charts are displayed in several sections, always with their reproducible code available. The gallery makes a focus on the tidyverse and [ggplot2](#). Feel free to suggest a chart or report a bug; any feedback is highly welcome. Stay in touch with the gallery by following it on [Twitter](#) or [Github](#). If you're new to R, consider following this [course](#).

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

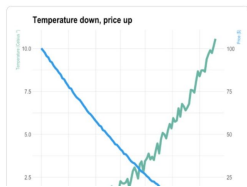
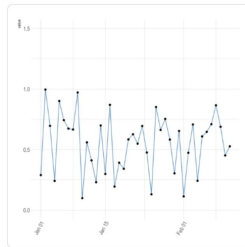
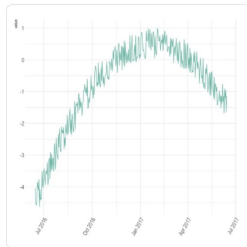
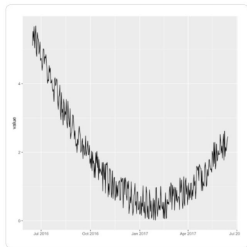
Shows the different types of charts in the collection

Times Series

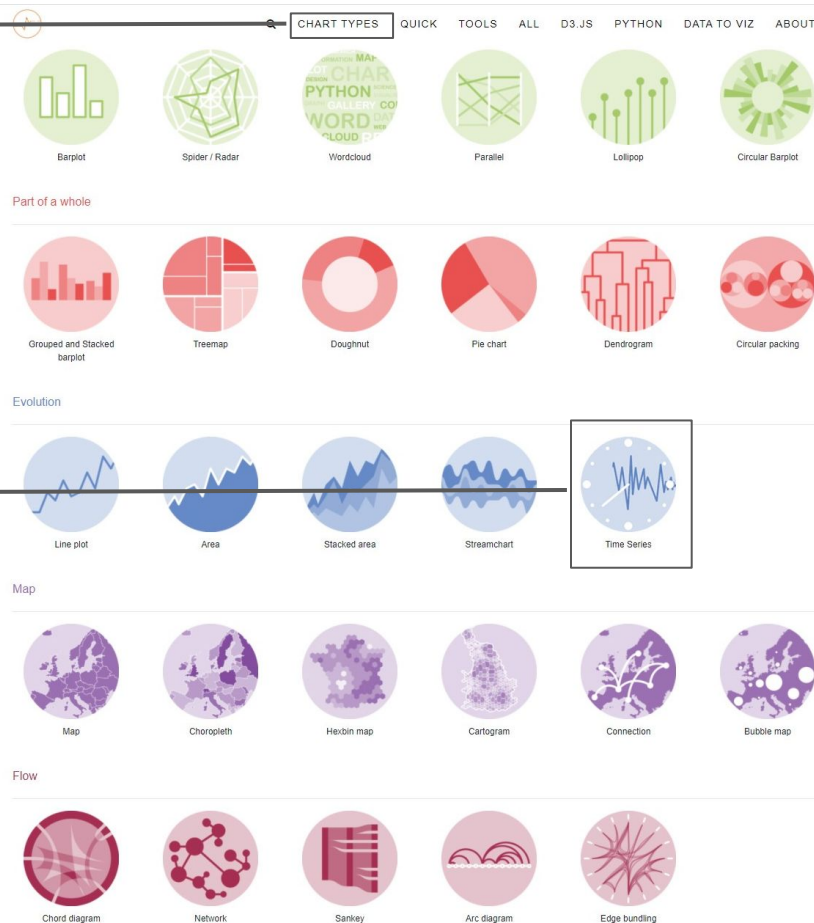
TIME SERIES WITH **GGPLOT2**

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

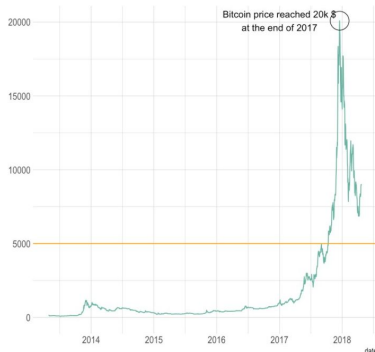
← the R Graph Gallery

Q CHART TYPES QUICK TOOLS ALL D3.JS PYTHON DATA TO VIZ ABOUT

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](#).



```
# 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=20000,
      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


Owner * **Repository name ***


 socapal / trade-method-presentation ✓

Great repository names are short and memorable. I

Description (optional)

Repository with the main example for the Method Presentation "Data Analyst Toolkit".

☒  **Public**
Anyone on the internet can see this repository. You choose who can commit.

☐  **Private**
You choose who can see and commit to this repository.


Initialize this repository with:
Skip this step if you're importing an existing repository.

☒ **Add a README file**
This is where you can write a long description for your project. [Learn more.](#)


☐ **Add .gitignore**
Choose which files not to track from a list of templates. [Learn more.](#)

☒ **Choose a license**
A license tells others what they can and can't do with your code. [Learn more.](#)

License: The Unlicense ▾




This will set  `main` as the default branch. Change the default name in your [settings](#).

[Create repository](#)


 socapal / trade-method-presentation


Unwatch ▾ 1 ☆ Star 0


<> Code ⓘ Issues 🔗 Pull requests ⚙️ Actions 📁 Projects 📖 Wiki 🛡️ Security 📊 Insights ⚙️ Settings


 main ▾  1 branch  0 tags

Go to file Add file ▾ [Code ▾](#)

 socapal Delete visualization.R 8f70dca 1 minute ago ⌚ 11 commits

 LICENSE Initial commit 21 hours ago

 README.md Update README.md 3 hours ago


README.md 


Trade 2021 Method Presentation Repository

Repository hosted at Github at [trade-method-presentation](#). Created by Karina Pérez Peña and Sebastián Ocampo Palacios.

About

Repository with the main example for the Method Presentation "Data Analyst Toolkit".

 Readme

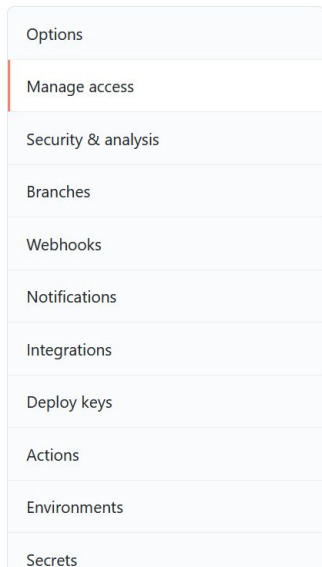
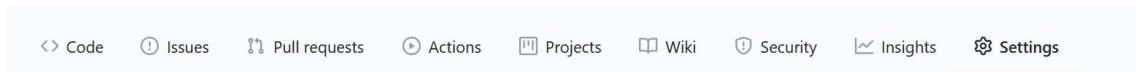
 Unlicense License

Releases

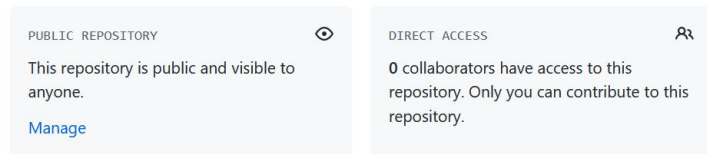
No releases published
[Create a new release](#)

Packages

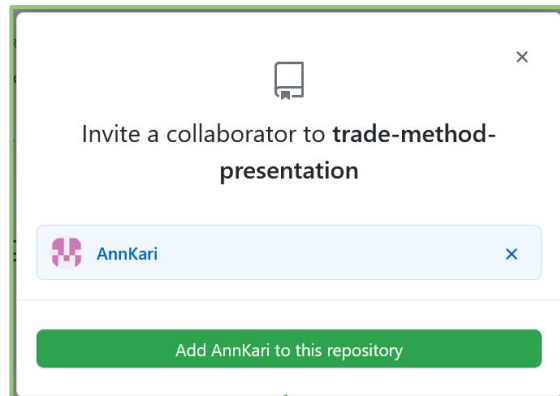
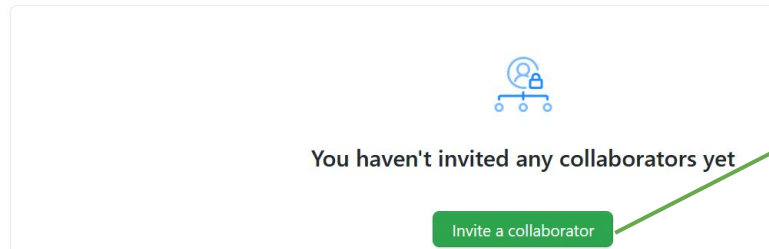
Step 3. Adding a Collaborator



Who has access



Manage access



Invitations can be accepted through the user's mail.

We are now ready to code collaboratively.

Step 4. Creating the code file

[Code](#) [Issues](#) [Pull requests](#) [Actions](#) [Projects](#) [Wiki](#) [Security](#) [Insights](#) [Settings](#)

main 1 branch 0 tags

Go to file

Add file

Code

socapal Delete visualization.R

Create new file

Upload files

11 commits

LICENSE	Initial commit	21 hours ago
README.md	Update README.md	3 hours ago

About

Repository with the main example for the Method Presentation "Data Analyst Toolkit".

Readme

Unlicense License

trade-method-presentation / Visualization.R in main

Cancel Changes

<> Edit new file Preview

```
1 Note: the gallery offers a section dedicated to line charts.
2
3 # Libraries
4 library(ggplot2)
5 library(dplyr)
6 library(plotly)
7 library(hrbrthemes)
8
9 # Load dataset from github
10 data <- read.table("https://raw.githubusercontent.com/holtzy/data_to_viz/master/Exemples/dataset.csv")
11 data$date <- as.Date(data$date)
12
13 # plot
14 data %>%
15   ggplot(aes(x=date, y=value)) +
16     geom_line(color="#69b3a2") +
17     ylim(0,22000) +
18     annotate(geom="text", x=as.Date("2017-01-01"), y=20089,
19            label="Bitcoin price reached 20k $\\nat the end of 2017") +
20     annotate(geom="point", x=as.Date("2017-12-17"), y=20089, size=10, shape=21, fill="orange") +
21     geom_hline(yintercept=5000, color="orange", size=.5) +
22     theme_ipsum()
```



Commit new file

Create Visualization.R

By creating this "Commit" we are modifying the main branch. On this box we should always comment what we have modified.
Created the code for visualization, we used Yan Holtz code.

- ☒ Commit directly to the `main` branch.
- ☐ Create a new branch for this commit and start a pull request. [Learn more about pull requests.](#)

Commit new file

Cancel

Step 5. Creating an issue to your collaborator

<> Code

! Issues 1

🔗 Pull requests

▶ Actions

📁 Projects

📖 Wiki

🛡 Security

...

Filters ▾

🔍 is:issue is:open

🏷 Labels 9

📅 Milestones 0

New issue

<> Code

! Issues

🔗 Pull requests

▶ Actions

📁 Projects

📖 Wiki

🛡 Security

...

Improve Visualization

Write

Preview

H B I ≡ <> 🔗 ≡ ≡ ☑ @ 🗨 ↶ ▾

@AnnKari could you remove the grid on the visualization and improve its colors?
Thank you!

Attach files by dragging & dropping, selecting or pasting them.

📄 Styling with Markdown is supported

Submit new issue

📄 Remember, contributions to this repository should follow our [GitHub Community Guidelines](#).

Assignees

👤 AnnKari

Labels

enhancement

good first issue

Projects

None yet

Milestone

No milestone

Linked pull requests

Successfully merging a pull request
may close this issue.

None yet

/

Pull requests Issues Marketplace Explore

Recent activity



Improve Visualization

enhancement

good first issue

socapal/trade-method-presentation · You opened this issue 11 minutes ago



Step 6. Implementing changes through a new branch



Search or jump to...



Pull requests

Issues

Marketplace

Explore



socapal / trade-method-presentation

Watch 1

Star 0

Fork 1

<> Code

Issues 1

Pull requests

Actions

Projects

Wiki

Security

Insights

main

1 branch

0 tags

Go to file

Add file

Code

Switch branches/tags



colors and grid

Branches

Tags

Create branch: colors and grid from 'main'

[View all branches](#)

94a538c 10 minutes ago

13 commits

Initial commit

22 hours ago

Update README.md

10 minutes ago

Create Visualization.R

22 minutes ago

About

Repository with the main example for the Method Presentation "Data Analyst Toolkit".

[Readme](#)

[Unlicense License](#)

Releases

color-and-grid

2 branches

0 tags

Go to file

Add file

Code

This branch is even with main.

[Pull request](#)

[Compare](#)

Step 7. Search questions, find answers.



Products

remove grid ggplot 2



Products

change colors ggplot 2

Home

PUBLIC

 Stack Overflow

Tags

Users

FIND A JOB

Jobs

Companies

TEAMS

What's this?

Remove grid, background color, and top and right borders from ggplot2

Ask Question

Asked 8 years, 9 months ago Active 11 months ago Viewed 211k times



108





40



I would like to reproduce the plot immediately below by using ggplot2. I can come close, but cannot remove the top and right borders. Below I present several attempts using ggplot2, including several suggestions found on or via Stackoverflow. Unfortunately I have not been able to get those suggestions to work.

I am hoping someone may be able to correct one or more of the code snippets below.

The Overflow Blog

-  Level Up: Mastering statistics with Python – part 2
-  What I wish I had known about single page applications

Featured on Meta

Step 8. Make changes & compare

<> Code 1 Issues 1 Pull requests Actions Projects Wiki Security Insights

Comparing changes

Choose two branches to see what's changed or to start a new pull request. If you need to, you can also [compare across forks](#).

base: main ← compare: color-and-grid ✓ Able to merge. These branches can be automatically merged.

Discuss and review the changes in this comparison with others. [Learn about pull requests](#)

Create pull request

1 commit

1 file changed

0 comments

1 contributor

Commits on Feb 27, 2021

Update Visualization.R

Showing 1 changed file with 2 additions and 2 deletions.

```
Visualization.R
@@ -13,10 +13,10 @@ data$date <- as.Date(data$date)
13 13 # plot
14 14 data %>%
15 15 ggplot(aes(x=date, y=value)) +
16 -   geom_line(color="#69b3a2") +
16 +   geom_line(color="black") +
17 17 ylim(0,22000) +
18 18 annotate(geom="text", x=as.Date("2017-01-01"), y=20089,
19 19         label="Bitcoin price reached 20k $(nat the end of 2017)") +
20 20 annotate(geom="point", x=as.Date("2017-12-17"), y=20089, size=10, shape=20)
21 21 geom_hline(yintercept=5000, color="orange", size=.5) +
22 -   theme_ipsum()
22 +   theme_ipsum(grid=FALSE)
```



Commit changes

Update Visualization.R

I have changed the color of the line and removed the grid leaving only horizontal lines

- ☒ Commit directly to the **color-and-grid** branch.
- ☐ Create a new branch for this commit and start a pull request. [Learn more about pull requests.](#)

Commit changes

Cancel

Step 9. Collaborator can now accept the pull request

The screenshot displays the GitHub interface for a pull request. At the top, navigation tabs include Code, Issues (1), Pull requests (1), Actions, Projects, Wiki, Security, Insights, and Settings. Below the navigation bar, a search bar shows 'is:pr is:open' and a 'New pull request' button is visible. On the left sidebar, a list of pull requests shows 'Update Visualization.R #2' as the selected item, with a green box highlighting it. The main content area shows the pull request details for 'Update Visualization.R #2'. It indicates that AnnKari wants to merge 1 commit into the 'main' branch from the 'color-and-grid' branch. A comment from AnnKari, posted 4 minutes ago, states: 'I have changed the color of the line and removed the grid leaving only horizontal lines'. Below the comment, a status bar shows 'Update Visualization.R' with a 'Verified' badge and a commit hash '9e498cd'. At the bottom, a green box highlights the status bar, which includes a message: 'Continuous integration has not been set up' (with a link to GitHub Actions) and 'This branch has no conflicts with the base branch' (with a note that merging can be performed automatically). A 'Merge pull request' button is visible at the bottom of the status bar. On the right side, the 'Reviewers' section shows 'socapal' as a reviewer, and the 'Assignees' section shows 'No one—assign yourself'.

<> Code ! Issues 1 🔗 Pull requests 1 ▶ Actions 📁 Projects 📖 Wiki 🛡 Security 📊 Insights ⚙ Settings

Filters ▾ 🔍 is:pr is:open 🏷 Labels 9 📌 Milestones 0 [New pull request](#)

☐ 🔗 1 Open ✓ 0 Closed

☐ 🔗 **Update Visualization.R**
#2 opened 2 minutes ago by AnnKari

Update Visualization.R #2

[Open](#) AnnKari wants to merge 1 commit into `main` from `color-and-grid`

🗨 Conversation 0 ↶ Commits 1 📋 Checks 0 📄 Files changed 1 +2 -2

AnnKari commented 4 minutes ago Collaborator 😊 ...

I have changed the color of the line and removed the grid leaving only horizontal lines

🔗 Update Visualization.R Verified 9e498cd

Add more commits by pushing to the `color-and-grid` branch on `socapal/trade-method-presentation`.

Continuous integration has not been set up
[GitHub Actions](#) and [several other apps](#) can be used to automatically catch bugs and enforce style.

✅ **This branch has no conflicts with the base branch**
Merging can be performed automatically.

[Merge pull request](#) You can also [open this in GitHub Desktop](#) or view [command line instructions](#).

Reviewers ⚙

Suggestions

socapal [Request](#)

Still in progress? [Convert to draft](#)

Assignees ⚙

No one—assign yourself

Labels ⚙

None yet

Projects ⚙

None yet

Milestone ⚙

No milestone

Step 10. Merge branches, delete the redundant branch

Update Visualization.R #2



AnnKari merged 1 commit into `main` from `color-and-grid` now



Conversation 0



Commits 1



Checks 0



Files changed 1



AnnKari commented 6 minutes ago

Collaborator



I have changed the color of the line and removed the grid leaving only horizontal lines



Update Visualization.R ...

Verified

9e498cd



AnnKari merged commit `3b96bd4` into `main` now

Revert



Pull request successfully merged and closed

You're all set—the `color-and-grid` branch can be safely deleted.

Delete branch

Step 11. Close Issue.

<> Code ! Issues 1 🔗 Pull requests ▶ Actions 📁 Projects 📖 Wiki 🛡 Security 📈 Insights ⚙ Settings

main ▾

1 branch

0 tags

Go to file

Add file ▾

Code ▾

About



Repository with the main example for the Method Presentation "Data Analyst Toolkit".

AnnKari Merge pull request #2 from socapal/color-and-grid ...

3b96bd4 1 minute ago ⌚ 17 commits

Improve Visualization #1

Open socapal opened this issue 34 minutes ago · 0 comments



socapal commented 34 minutes ago

Owner ⚙ ...

@AnnKari could you remove the grid on the visualization and improve its colors? Thank you!



socapal added enhancement good first issue labels 34 minutes ago



socapal assigned AnnKari 34 minutes ago



Write Preview

H B I ☰ <> 🔗 ☰ ☰ ☑ @ ↩ ↶

Leave a comment

Attach files by dragging & dropping, selecting or pasting them.



Close issue

Comment

Improve Visualization #1

Edit New issue

Closed socapal opened this issue 39 minutes ago · 0 comments



socapal commented 39 minutes ago

Owner ⚙ ...

Assignees
AnnKari

Labels
enhancement good first issue

Projects
None yet

Milestone
No milestone



socapal added enhancement good first issue labels 39 minutes ago



socapal assigned AnnKari 39 minutes ago

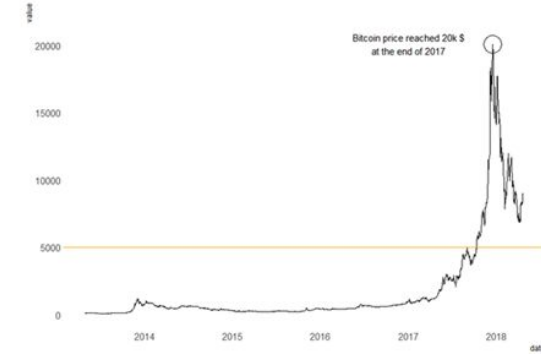


AnnKari closed this 3 minutes ago

Finally, let's publish our graph!

Step 12. Publication

r/dataisbeautiful · Posted by [redacted] 1 minutes ago
[OC] Bitcoin price from 2014 to mid 2018



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. Plagiarism 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 look 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 deflated that currency!

Vote Reply Give Award Share Report Save

Bitcoin price 2014-2018.

#dataviz #DataMoney



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!

Repository Access

Repository created by Anna Karina Pérez Peña and Sebastián Ocampo Palacios. It can be accessed at the following link: <https://github.com/socapal/trade-method-presentation>.

There you will find more useful resources to learn Github.

Last updated on March 4, 2021.

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>.

r/dataisbeautiful. Reddit, hosted at <https://www.reddit.com/r/dataisbeautiful/>.