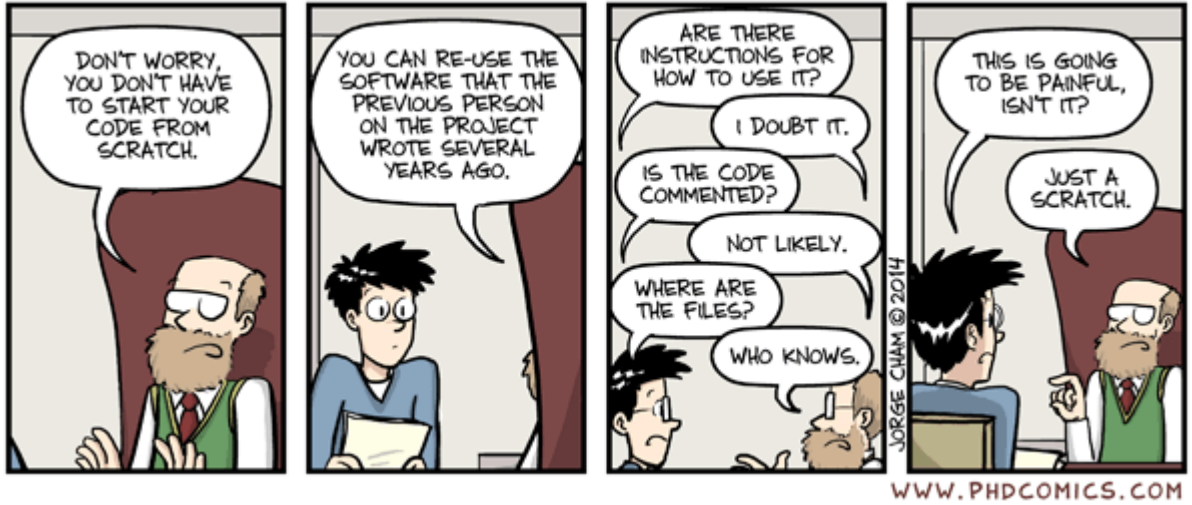
Jak mogę sprawić, by mój kod był bardziej powtarzalny dla mnie i innych?

Powtarzalność jest trudna…

Powtarzalność jest trudna…



"Scratch" Jorge Cham (phdcomics.com)

Powtarzalność badań z workflowr

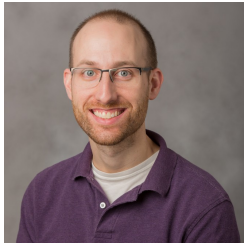
Narzędzie do organizowania, powtarzalności, wersjonowania i udostępniania projektów analizy danych

Opracowany przez ….

Opracowany przez ….

Matthew Stephens
University of Chicago

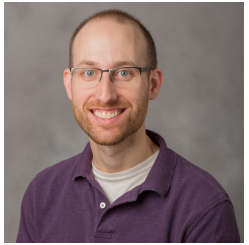
Opracowany przez ….



Matthew Stephens
University of Chicago

John Blischak
University of Chicago

Opracowany przez ….



Matthew Stephens
University of Chicago

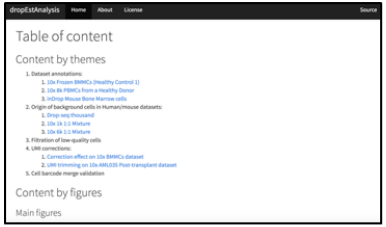
John Blischak
University of Chicago

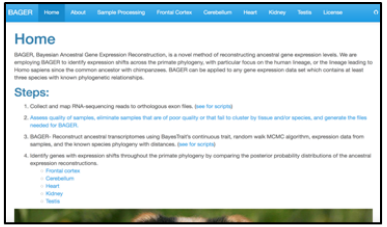
Peter Carbonetto
University of Chicago

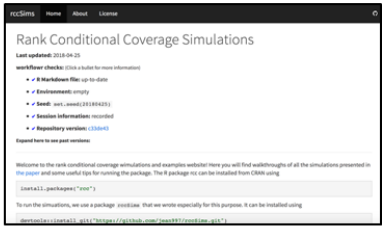
THE UNIVERSITY OF CHICAGO

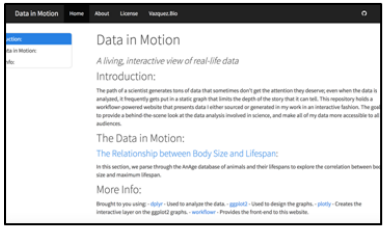
workflowr

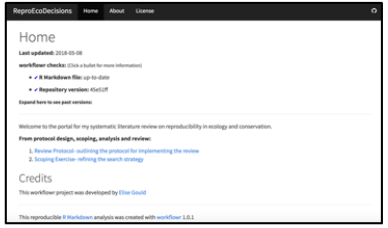
- Zorganizowany
- Powtarzalny
- Wersjonowany
- Dostępny

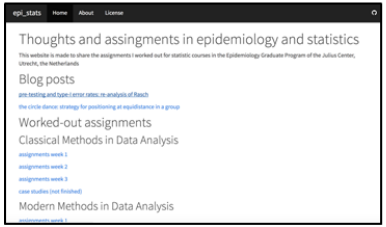


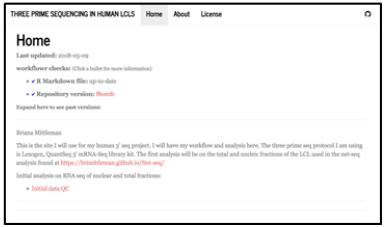












Zorganizowany

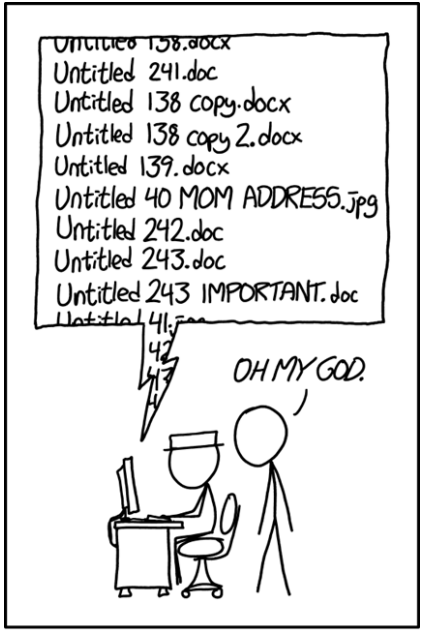
Zorganizowana kuchnia

Zorganizowana kuchnia

Zorganizowana kuchnia

Źródło: homebook, shutter stock

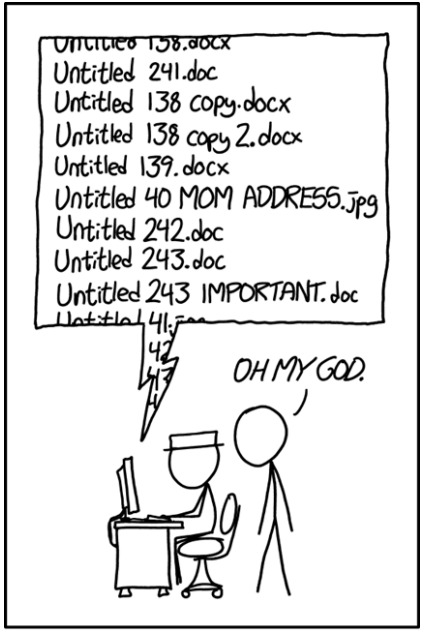
Zorganizowany projekt…



PROTIP: NEVER LOOK IN SOMEONE ELSE'S DOCUMENTS FOLDER.

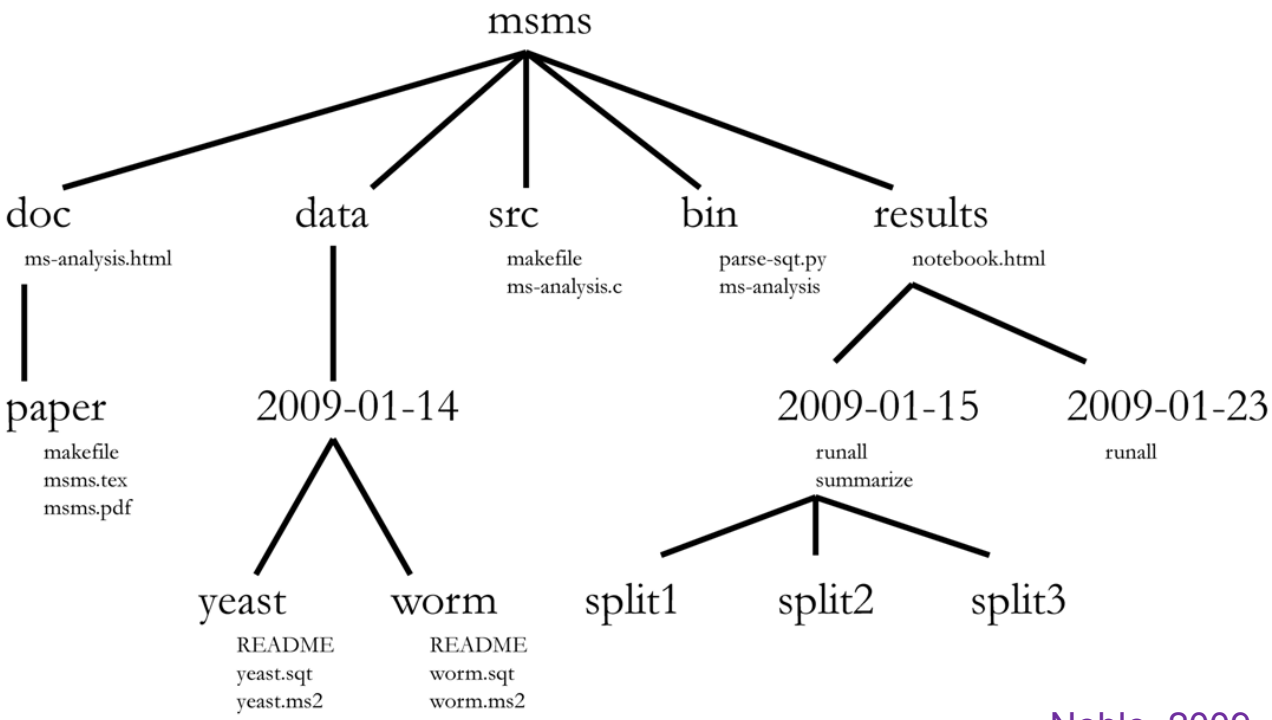
xkcd.com/1459/

Zorganizowany projekt…



PROTIP: NEVER LOOK IN SOMEONE ELSE'S DOCUMENTS FOLDER.

xkcd.com/1459/



Noble, 2009

Zorganizowana struktura katalogów

Zorganizowana struktura katalogów

```
wflow_start("myproject")
```

Zorganizowana struktura katalogów

```
wflow_start("myproject")
```

```
myproject/
├── analysis
│   ├── about.Rmd
│   ├── index.Rmd
│   ├── license.Rmd
│   └── _site.yml
├── code
│   └── README.md
├── data
│   └── README.md
├── docs
├── myproject.Rproj
├── output
│   └── README.md
├── README.md
└── _workflowr.yml
```

Powtarzalny

Definicja powtarzalności

Definicja powtarzalności

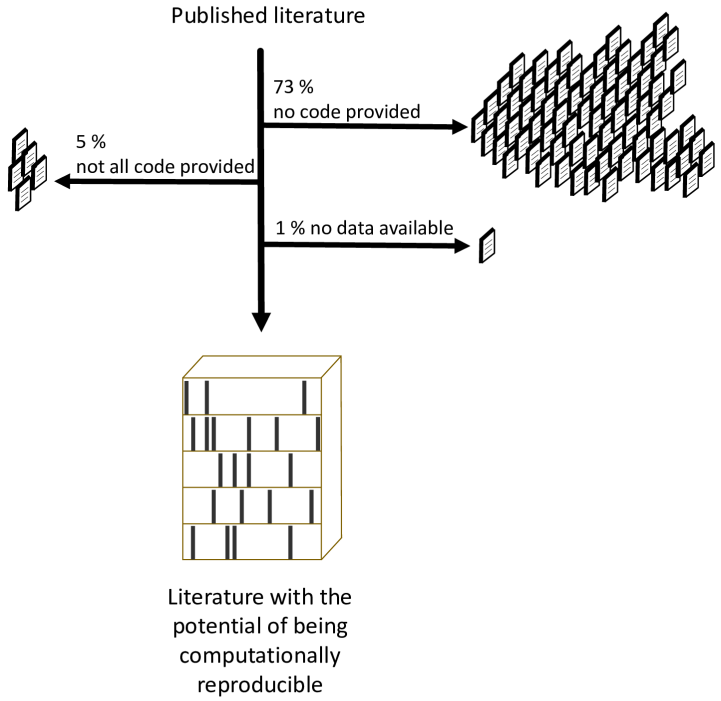
Inna osoba* jest w stanie uruchomić ten sam kod na tych samych danych i otrzyma te same wyniki

Definicja powtarzalności

Inna osoba* jest w stanie uruchomić ten sam kod na tych samych danych i otrzyma te same wyniki

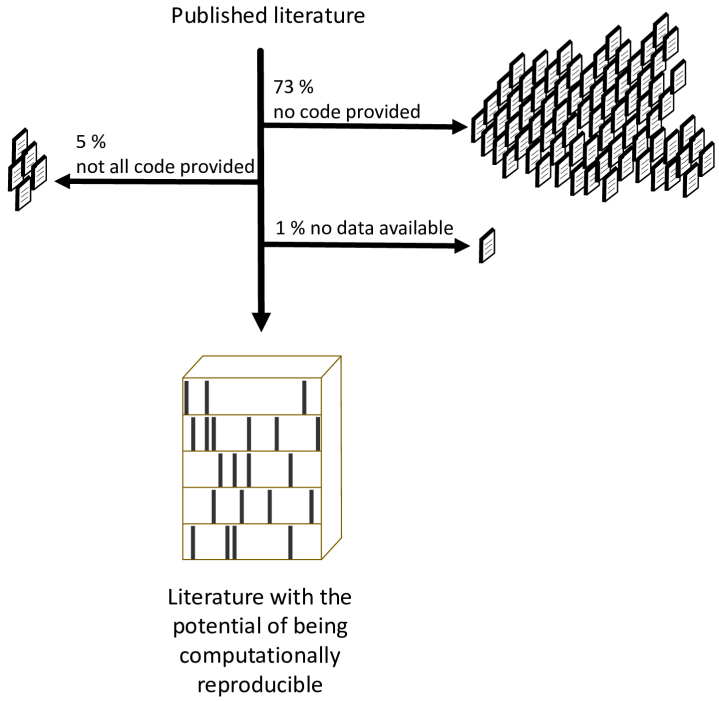
*może nią również być każdy z nas za 6 miesięcy, gdy nie odpowiada na maile bo nie wie, gdzie co było…

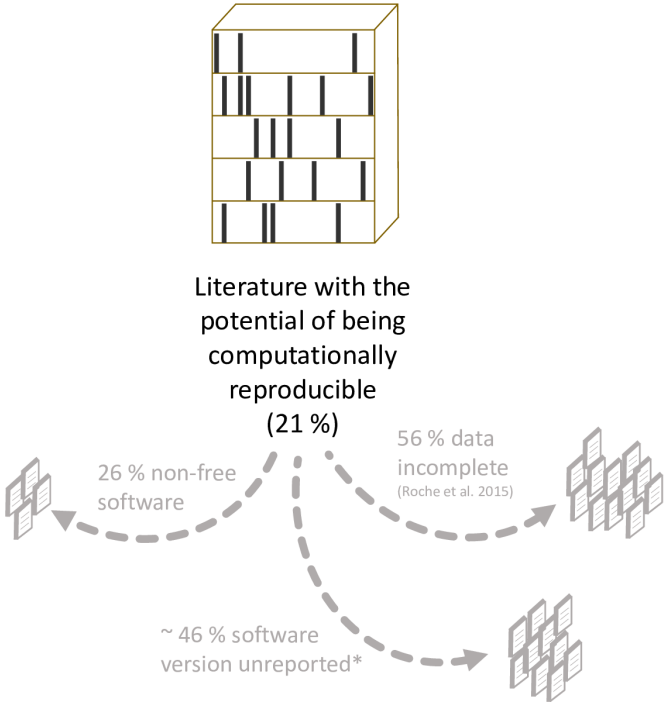
Powtarzalność badań w ekologii w 2020 roku (365 prac, 4 lata)



Culina et al., 2020

Powtarzalność badań w ekologii w 2020 roku (365 prac, 4 lata)

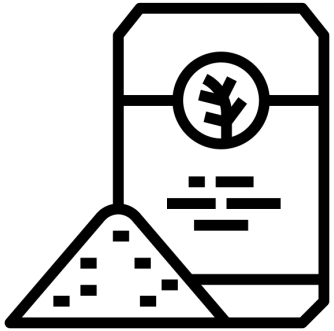




Culina et al., 2020

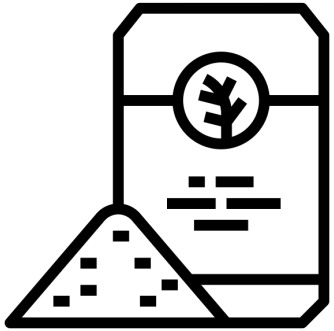
Pieczenie

Pieczenie

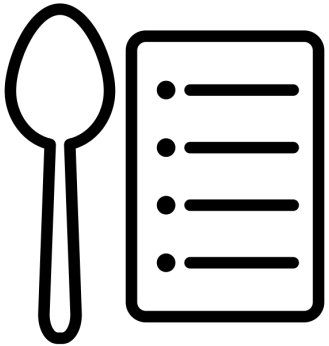


Surowe składniki

Pieczenie

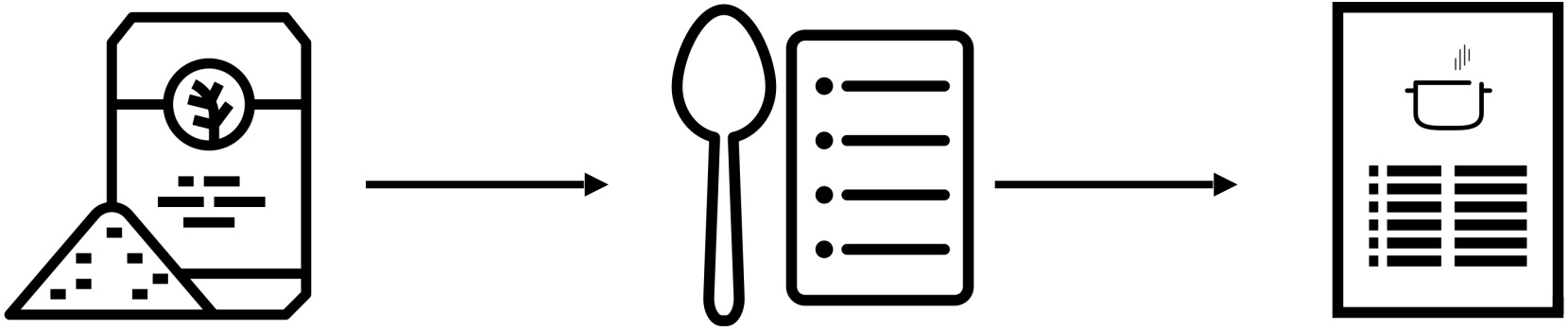


Surowe składniki



Seria kroków

Pieczenie



Surowe składniki

Seria kroków

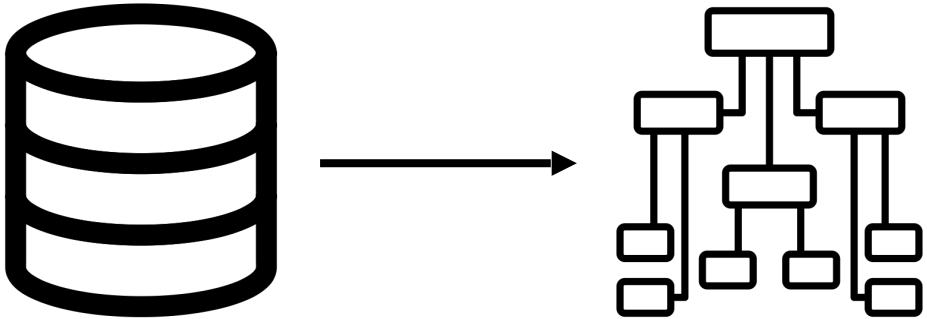
Opublikowany przepis

Analiza danych

Analiza danych

Surowe dane

Analiza danych

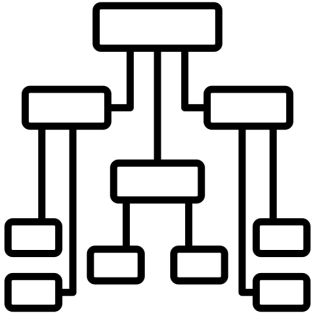


Surowe dane

Seria kroków

Analiza danych

Surowe dane



Seria kroków

Opublikowane wyniki

Tworzenie strony internetowej

Tworzenie strony internetowej

```
wflow_build()
```

Tworzenie strony internetowej

```
wflow_build()
```

```
myproject/
├── analysis
│   ├── about.Rmd
│   ├── index.Rmd
│   ├── license.Rmd
│   └── _site.yml
├── code
│   └── README.md
├── data
│   └── README.md
├── docs
├── myproject.Rproj
├── output
│   └── README.md
├── README.md
└── _workflowr.yml
```

Używaj względnych instrukcji

Używaj względnych instrukcji

- Pieczemy w piekarniku na czwartej półce

Używaj względnych instrukcji

- Pieczemy w piekarniku na czwartej półce

- Pieczemy w piekarniku na <u>najniższej</u> półce

Używaj względnych ścieżek do plików

Używaj względnych ścieżek do plików

```
dane<-read.table('/Users/joanna/Desktop/single_cell/cell.txt')
```

Używaj względnych ścieżek do plików

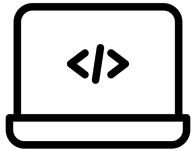
```
dane<-read.table('/Users/joanna/Desktop/single_cell/cell.txt')
```

Używaj względnych ścieżek do plików

```
dane<-read.table('/Users/joanna/Desktop/single_cell/cell.txt')
```

```
dane<-read.table('data/cell.txt')
```

Używaj względnych ścieżek do plików



```
dane<-read.table('/Users/joanna/Desktop/single_cell/cell.txt')
```

```
dane<-read.table('data/cell.txt')
```

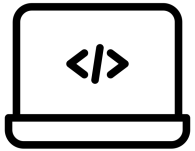
Radzenie sobie z losowością

- Dodaj szczyptę soli

Radzenie sobie z losowością

- Dodaj szczyptę soli

- Podaj 100 liczb losowych wybranych ze standardowego rozkładu normalnego



Ustawienie seed'a

```
> set.seed(1)
> x <- rnorm(100)
> head(x, 3)
[1] -0.6264538  0.1836433 -0.8356286
> summary(x)
Min. 1st Qu.  Median    Mean 3rd Qu.   Max.
-2.2147 -0.4942  0.1139  0.1089  0.6915  2.4016
```

Workflowr automatycznie rozpoznaje set.seed

workflowr

✔ **Seed:** `set.seed(1)`

The command `set.seed(1)` was run prior to running the code in the R Markdown file. Setting a seed ensures that any results that rely on randomness, e.g. subsampling or permutations, are reproducible.

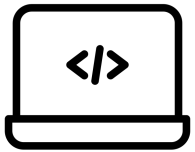
Zapisz narzędzia i składniki

- Rodzaje garnków i patelni
- Nazwy firm przypraw
- Daty przydatności do spożycia
- ...

Zapisz narzędzia i składniki

- Rodzaje garnków i patelni
- Nazwy firm przypraw
- Daty przydatności do spożycia
- ...

- System operacyjny
- Wersja R
- Wersja pakietów R
- ...



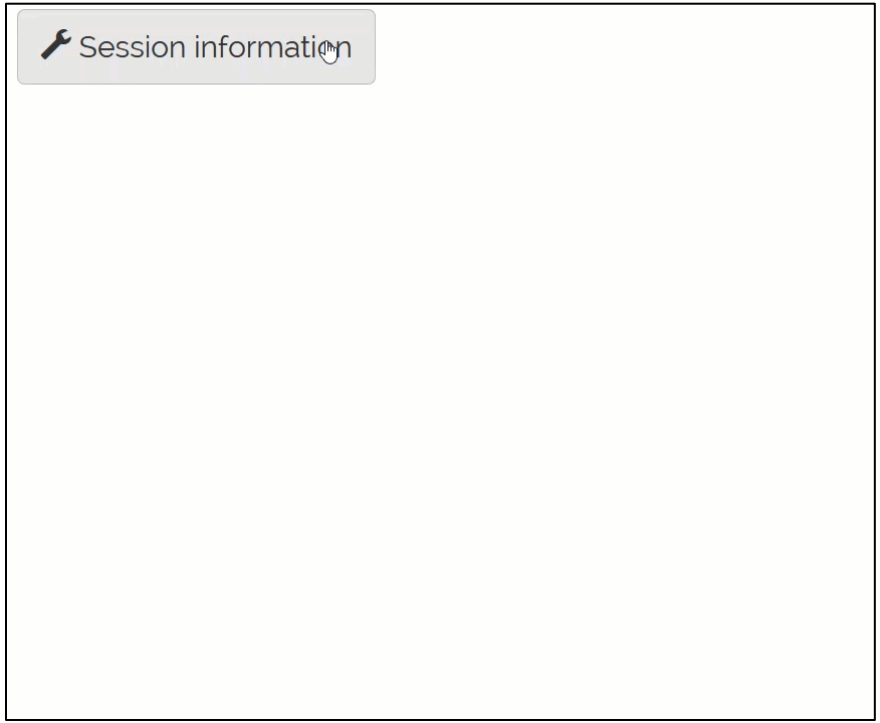
Informacja o sesji

```
> sessionInfo()
R version 4.0.2 (2020-06-22)
Platform: x86_64-w64-mingw32/x64 (64-bit)
Running under: Windows 10 x64 (build 19041)

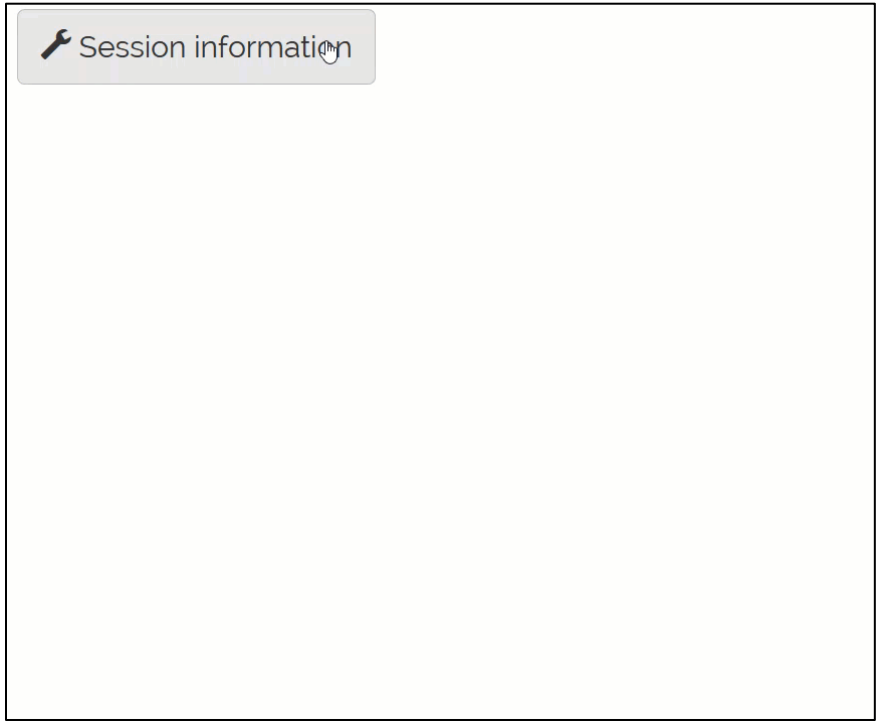
<some content omitted for brevity>

loaded via a namespace (and not attached):
  [1] compiler_4.0.2  htmltools_0.5.0 tools_4.0.2
  [4] yaml_2.2.1    rmarkdown_2.3   knitr_1.29
  [7] xfun_0.15     digest_0.6.25   rlang_0.4.6
  [10] evaluate_0.14
```

Workflowr zapamiętuje informacje o sesji…

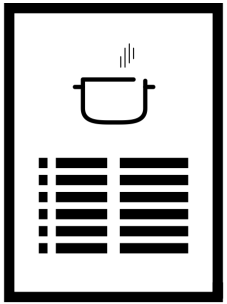


Workflowr zapamiętuje informacje o sesji…

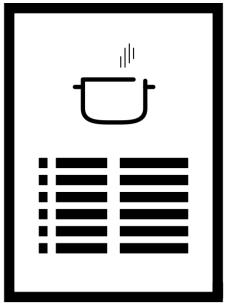


Wersjonowany

Zapisywanie zmian

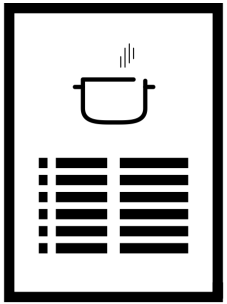


Zapisywanie zmian



2022-12-05: Zwiększam czas pieczenia.

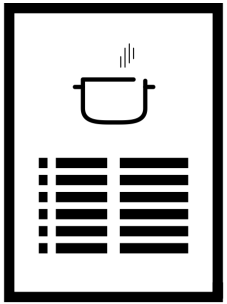
Zapisywanie zmian



2022-12-15: Zastępuję margarynę masłem.

2022-12-05: Zwiększam czas pieczenia.

Zapisywanie zmian



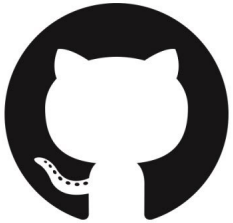
2022-12-20: Dodaję więcej cukru. Nie jest zbyt słodkie.

2022-12-15: Zastępuję margarynę masłem.

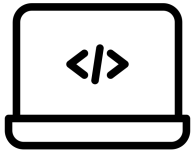
2022-12-05: Zwiększam czas pieczenia.

Co to jest GitHub?

mojewypieki



GitHub



Kontrola wersji (Git)

History for **zajecia** / **data**

Commits on Jan 16, 2023

Usunięcie kod.R z folderu data
zjanna committed 2 minutes ago — 0639586

Dodanie kodu
zjanna committed 15 minutes ago — 6fbfaf2

Commits on Jan 11, 2023

Add files
zjanna committed 5 days ago — f3745cb

Start workflowr project.
zjanna committed 5 days ago — e4c321e

End of commit history for this file

Newer Older

Zestawienie poprzednich wersji

Zajęcia Home About License

Home

workflowr

Zajęcia z R

Poniżej znajdują się materiały do zajęć:

- wizualizacje
- wprowadzenie

Zestawienie poprzednich wersji

Zajęcia Home About License

Home

workflowr

Zajęcia z R

Poniżej znajdują się materiały do zajęć:

- wizualizacje
- wprowadzenie

Publikowanie

Blischak et al. 2019

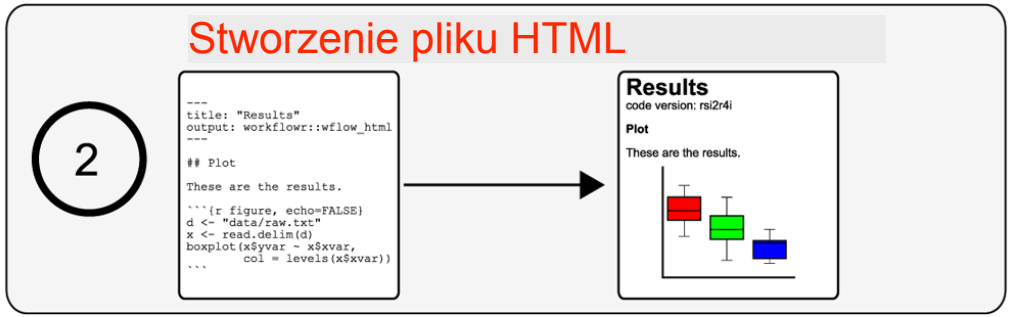
```
> wflow_publish("analysis/results.Rmd")
```

Publikowanie

Blischak et al. 2019

```
> wflow_publish("analysis/results.Rmd")
```

Publikowanie



Blischak et al. 2019

```
> wflow_publish("analysis/results.Rmd")
```

Publikowanie

Zatwierdzenie zmian w pliku Rmd

1

```
---
title: "Results"
output: workflowr::wflow_html
---

## Plot

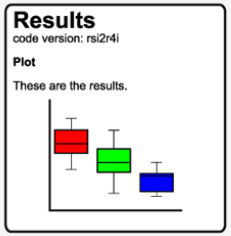
These are the results.

```{r figure, echo=FALSE}
d <- "data/raw.txt"
x <- read.delim(d)
boxplot(x$yvar ~ x$xvar,
        col = levels(x$xvar))
```
```

Stworzenie pliku HTML

2

```
---
title: "Results"
output: workflowr::wflow_html
---

## Plot

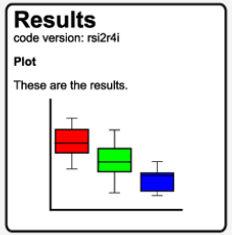
These are the results.

```{r figure, echo=FALSE}
d <- "data/raw.txt"
x <- read.delim(d)
boxplot(x$yvar ~ x$xvar,
        col = levels(x$xvar))
```
```



Zatwierdzenie zmian w pliku HTML

3



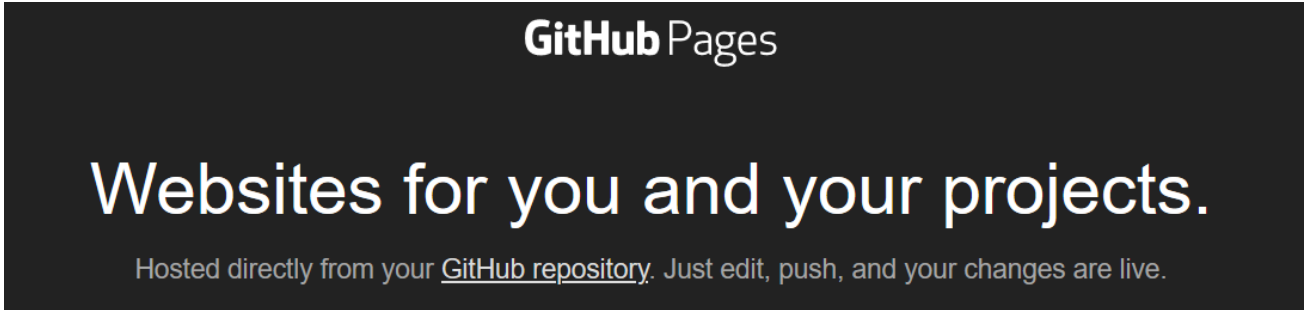
Blischak et al. 2019

Dostępny

Dzielenie się kodem online

Dzielenie się kodem online

Dzielenie się kodem online



https://pages.github.com/

workflowr umożliwia udostępnianie projektu

workflowr umożliwia udostępnianie projektu

Wstępne ustawienia:

```
wflow_use_github("username")
```

workflowr umożliwia udostępnianie projektu

Wstępne ustawienia:

```
wflow_use_github("username")
```

Udostępnienie najnowszych zmian:

```
wflow_git_push()
```

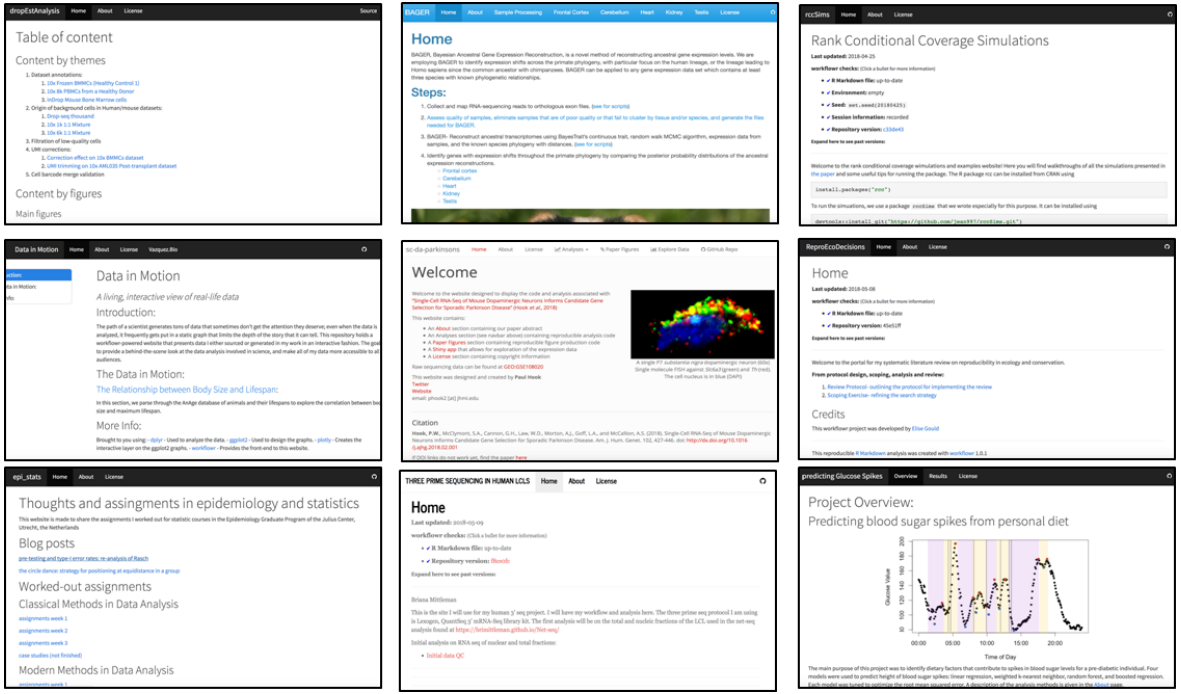
workflowr

Zorganizowany

Powtarzalny

Wersjonowany

Dostępny



Workflowr w praktyce…

1. `install.packages("workflowr")`
2. Załóż konto na GitHub
3. Wykonuj następujące kroki:

1. Załadowanie pakietu workflowr

```
library(workflowr)
```

2. Konfiguracja gita (krok ten może być wykonany tylko raz na danym komputerze)

```
# Replace the example text with your information
wflow_git_config(user.name = "Your Name",
                 user.email = "email@domain")
```

3. Rozpoczęcie projektu

Uruchomienie poniższego kodu spowoduje stworzenie kompletnej struktury katalogów workflowr:

```
wflow_start(directory = ".",
            name = "zajecia",
            existing = TRUE)
```

Najważniejsze katalogi to `analysis/`, który to zawiera pliki Rmd, oraz `docs/`, który zawiera pliki html.

4. Stworzenie strony internetowej

Tworzenie strony internetowej z domyślnych plików Rmd znajdujących się w katalogu `analysis/`:

```
wflow_build()
```

Strona automatycznie się otworzy.

5. Dodawanie nowych analiz

Przeniesienie naszych analiz do projektu workflowr:

```
file.rename(„wprowadzenie.Rmd”, „analysis/wprowadzenie.Rmd")
file.rename(„wizualizacje.Rmd”, „analysis/wizualizacje.Rmd")
file.rename(„studenci.xlsx”, „data/studenci.xlsx")
```

Przenosimy pliki `wprowadzenie.Rmd` oraz `wizualizacje.Rmd` do folderu `analysis/` oraz dane `studenci.xlsx` do folderu `data/`.

6. Tworzenie nowej analizy

```
wflow_build()
```

Domyślnie `wflow_build()` tworzy nowe albo uaktualnia modyfikacje z plików Rmd.

7. Dodawanie linków do nowych analiz

Otwórz `analysis/index.Rmd` i dodaj następującą komendę:

```
[wizualizacje](wizualizacje.html)
[wprowadzenie](wprowadzenie.html)
```

To jest składnia Markdown dla stworzenia hiperlinku ze strony głównej do pliku HTML z analizami.

Następnie wprowadź kod `wflow_build()` w celu wprowadzenia uaktualnień:

```
wflow_build()
```

8. Opublikowanie strony internetowej

Stworzona nowa analiza może zostać opublikowana poprzez:

```
wflow_publish(c("analysis/*Rmd", „data/studenci.xlsx”,
„docs/*html),message = „Dodanie nowych analiz”)
```

Opublikowanie pliku Rmd spowoduje uaktualnienie nie tylko wersji Rmd ale również pliku HTML.

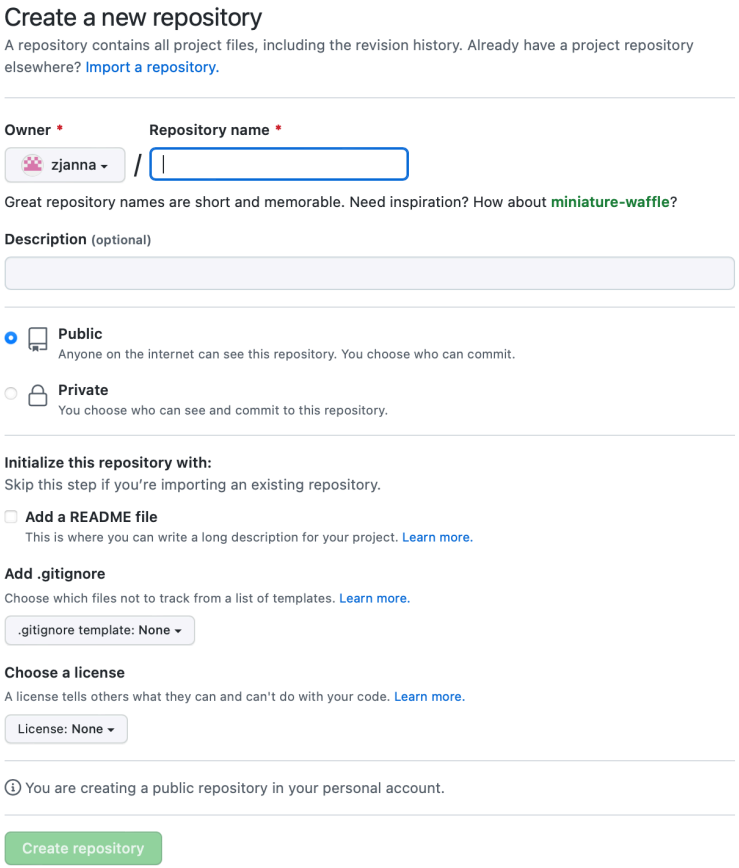
9. Stworzenie repozytorium na GitHub

9. Stworzenie repozytorium na GitHub

Idź do https://github.com/new/ i nazwij nowe repozytorium “zajecia”

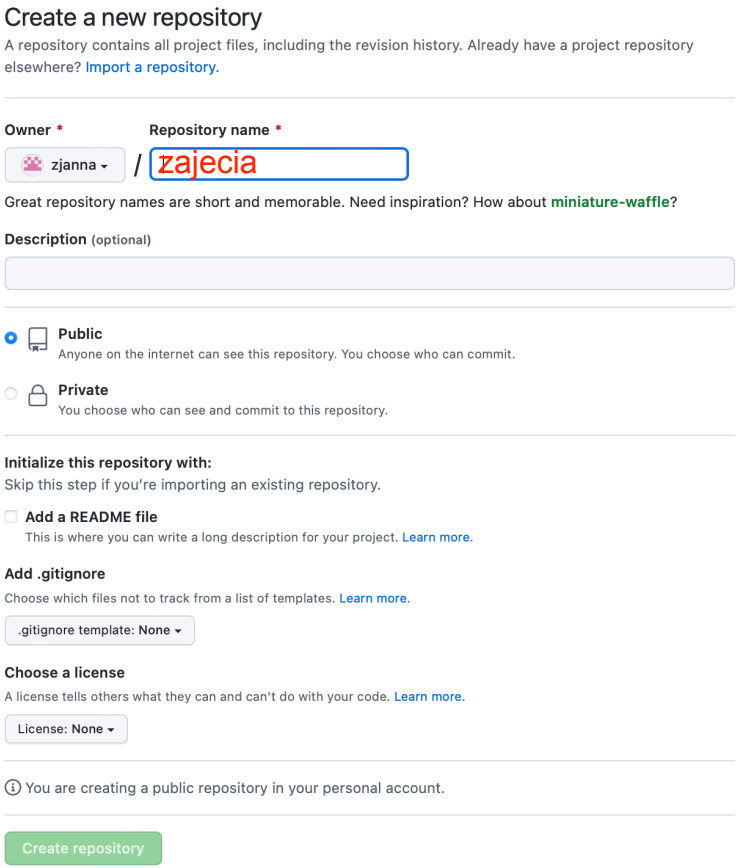
9. Stworzenie repozytorium na

Idź do https://github.com/new/ i nazwij nowe repozytorium “zajecia”



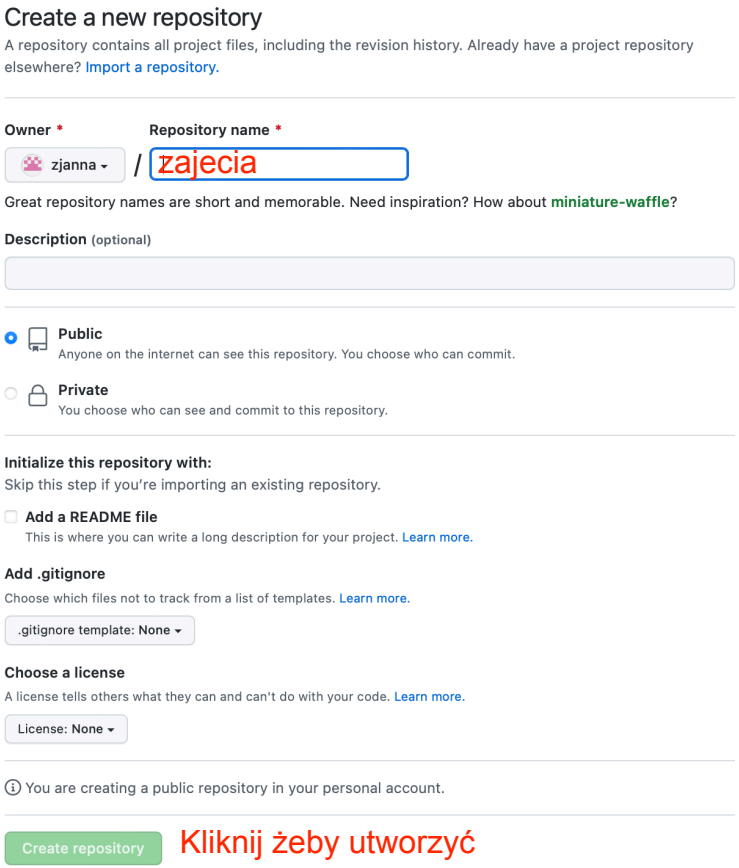
9. Stworzenie repozytorium na

Idź do https://github.com/new/ i nazwij nowe repozytorium “zajecia”



9. Stworzenie repozytorium na

Idź do https://github.com/new/ i nazwij nowe repozytorium “zajecia”



10. Połączenie repozytorium GitHub z projektem workflowr

Zamień "<github-username>" swoją nazwą użytkownika z GitHub, np.
user = „zjanna"

```
wflow_git_remote("origin",
                 user = "<github-username>",
                 repo = „zajecia")
```

11. Wyślij projekt workflowr do GitHub

11. Wyślij projekt workflowr do GitHub

```
wflow_git_push()
```

11. Wyślij projekt workflowr do GitHub

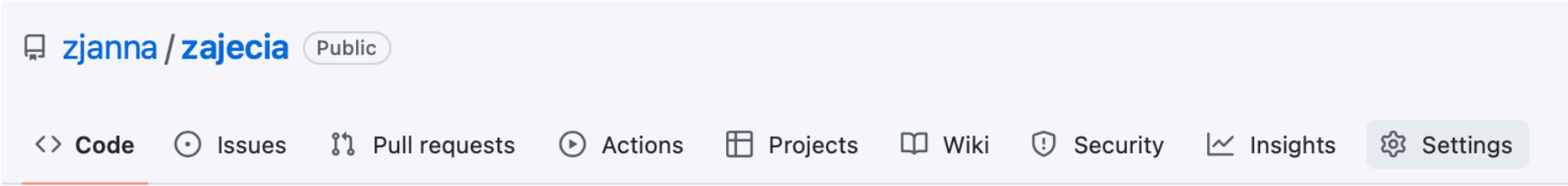
```
wflow_git_push()
```

12. Zmień ustawienia projektu

11. Wyślij projekt workflowr do GitHub

```
wflow_git_push()
```

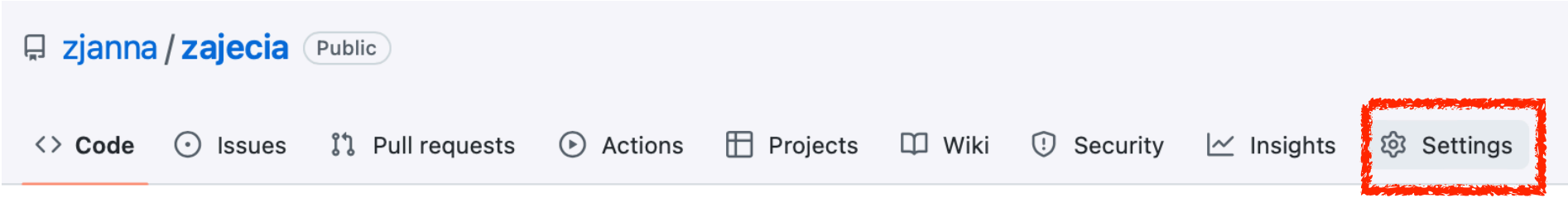
12. Zmień ustawienia projektu



11. Wyślij projekt workflowr do GitHub

```
wflow_git_push()
```

12. Zmień ustawienia projektu



11.

12.

wfl

zjann

Code

General

Access

Collaborators

Moderation options

Code and automation

Branches

Tags

Actions

Webhooks

Environments

Codespaces

Pages

Security

Code security and analysis

Deploy keys

Secrets and variables

Integrations

GitHub apps

Email notifications

GitHub Pages

GitHub Pages is designed to host your personal, organization, or project pages from a GitHub repository.

Your site is live at https://zjanna.github.io/zajecia/
Last deployed by zjanna 4 days ago

Visit site

Build and deployment

Source

Deploy from a branch

Branch
Your GitHub Pages site is currently being built from the `/docs` folder in the `master` branch. Learn more.

master /docs Save

Learn how to add

Select folder
/ (root)
/docs

Your site was last the pages build and deployment workflow.
Learn more about kflows

Custom domain

Custom domain
Custom domains allow you to serve your site from a domain other than `zjanna.github.io`. Learn more.

Save Remove

DNS records should point to the internationalized domain name .

Enforce HTTPS
— Required for your site because you are using the default domain (`zjanna.github.io`)

HTTPS provides a layer of encryption that prevents others from snooping on or tampering with traffic to your site.
When HTTPS is enforced, your site will only be served over HTTPS. Learn more.

Settings

11.

12.

wfl

zjann

Code

General

Access

Collaborators

Moderation options

Code and automation

Branches

Tags

Actions

Webhooks

Environments

Codespaces

Pages

Security

Code security and analysis

Deploy keys

Secrets and variables

Integrations

GitHub apps

Email notifications

GitHub Pages

GitHub Pages is designed to host your personal, organization, or project pages from a GitHub repository.

Your site is live at https://zjanna.github.io/zajecia/
Last deployed by zjanna 4 days ago

Visit site

Build and deployment

Source

Deploy from a branch

Branch
Your GitHub Pages site is currently being built from the `/docs` folder in the `master` branch. Learn more.

master /docs Save

Learn how to add

Select folder
/ (root)
/docs

Your site was last the pages build and deployment workflow.
Learn more about kflows

Custom domain

Custom domain
Custom domains allow you to serve your site from a domain other than `zjanna.github.io`. Learn more.

Save Remove

DNS records should point to the internationalized domain name .

Enforce HTTPS
— Required for your site because you are using the default domain (`zjanna.github.io`)

HTTPS provides a layer of encryption that prevents others from snooping on or tampering with traffic to your site.
When HTTPS is enforced, your site will only be served over HTTPS. Learn more.

Settings

13. Link do strony…

https://<github-username>.github.io/zajecia/

13. Link do pomocy

https://workflowr.github.io/workflowr/reference/index.html

Wykorzystane materiały

2021 Toronto Workshop on Reproducibility: Facilitating reproducible and open research with workflowr and workflowr.io" by John Blischak, https://bit.ly/workflowr-toronto

"Create and share reproducible code with R Markdown and workflowr" by John Blischak and Peter Carbonetto

"Why R? Webinar: Reproducible research with workflowr" by John Blischak, https://bit.ly/why-r-workflowr

Noble WS. A quick guide to organizing computational biology projects. PLoS Comput Biol. 2009 Jul;5(7):e1000424. doi: 10.1371/journal.pcbi.1000424. Epub 2009 Jul 31. PMID: 19649301; PMCID: PMC2709440.

Culina A, van den Berg I, Evans S, Sánchez-Tójar A (2020) Low availability of code in ecology: A call for urgent action. PLOS Biology 18(7).