SQL-Datenbanken abfragen und Versionierung mit `git`

Unit 6

BASEL LANDSCHAFT
FINANZ- UND KIRCHENDIREKTION
AMT FÜR DATEN UND STATISTIK

Ziele für heute

1. eine Verbindung zu einer Datenbank herstellen und SQL-Abfragen mit `{DBI}` ausführen
2. den grundlegenden Workflow von Git in RStudio erklären
3. die wichtigsten Konzepte des Kurses zusammenfassen

Praktikum 06a

CO2-Emissionen

30:00


essential skill
COURSE CONTENT
ANOTHER IMPORTANT THING
ABSOLUTELY CRITICAL
CORE
@allison_horst

SQL Database

SQL Database

- `DBI` + `RPostgres` → Verbindung mit SQL Datenbank
- dbplyr → `dplyr`-Verben mit Datenbank

```
1 library(DBI) # dbConnect()
2 library(RPostgres) # Postgres()
```

SQL Database

- `DBI` + `RPostgres` → Verbindung mit SQL Datenbank
- dbplyr → `dplyr`-Verben mit Datenbank

```r
library(DBI) # dbConnect()
library(RPostgres) # Postgres()

con <- dbConnect(
  Postgres(),
  sslmode = "allow",
  host = "host-name",
  port = port-number,
  dbname = "db-name",
  user = "username",
  password = "my_password"
)
```

⛔ Niemals Passwort im Skript speichern!

SQL Database

- `DBI` + `RPostgres` → Verbindung mit SQL Datenbank
- dbplyr → `dplyr`-Verben mit Datenbank

```
1  library(DBI) # dbConnect()
2  library(RPostgres) # Postgres()
3
4  con <- dbConnect(
5    Postgres(),
6    sslmode = "allow",
7    host = "host-name",
8    port = port-number,
9    dbname = "db-name",
10   user = "username",
11   password = rstudioapi::askForPassword("Datenbank Passwort")
12 )
```

config.yml

```yaml
default:
  DWH_var:
    host: "host-name"
    user: "username"
    password: "DWH_pass"
    port: port-number
    dbname: "db-name"
```

```r
library(DBI) # dbConnect()
library(RPostgres) # Postgres()

con <- dbConnect(
  Postgres(),
  sslmode = "allow",
  host = config::get("DWH_var")$host,
  dbname = config::get("DWH_var")$dbname,
  user = config::get("DWH_var")$user,
  password = config::get("DWH_var")$password,
  port = config::get("DWH_var")$port
)
```

config.yml

```
default:
  DWH_var:
    host: "host-name1"
    user: "username"
    password: "DWH_pass"
    port: port-number1
    dbname: "db-name1"
  DWH_hrm:
    host: "host-name"
    user: "username"
    password: "DWH_pass"
    port: port-number2
    dbname: "db-name2"

andere_DB:
  DB_fin:
    host: "host-name3"
    user: "username"
```

SQL Daten Bearbeiten

```
1 con
```

```
1 dbListObjects(con)
```

SQL Daten Bearbeiten

```
1  dbListObjects(con, Id(schema = "schema_name"))
```

SQL Daten Bearbeiten

```
1 library(tidyverse)
2 df_db <- tbl(con, Id(schema = "schema_name", table = "table_name"))
3 class(df_db)
```

```
1 glimpse(df_db, width = 20)
```

SQL Daten Bearbeiten

```r
df_db_small <- df_db |>
  select(var1, district_k, var2) |>
  mutate(
    var1 = as.character(var1), # dbplyr cannot translate factor() to sql
    district_k = as.character(district_k),
    district_k = case_when(
      district_k == "1301" ~ "Arlesheim",
      district_k == "1302" ~ "Laufen",
      district_k == "1303" ~ "Liestal",
      district_k == "1304" ~ "Sissach",
      .default = "Waldenburg"
    )
  ) |>
  summarise(var3 = sum(var2), .by = c(var1, district_k))
```

SQL Daten Bearbeiten

```
df_db_small |>
  ggplot(
    aes(
      x = var1,
      y = var3,
      fill = fct_reorder(
        district_k,
        var3
      )
    )
  ) +
  geom_col() +
  labs(
    x = "",
    y = "Total",
    fill = "Bezirk"
  ) +
  scale_y_continuous(
```

SQL Query: `show_query()`

```
1 df_db_small |>
2   show_query()
```

SQL Queries in Quarto Dateien

````
```{sql}
#| connection: con
#| output.var: "db_sql_small"
#| code-line-numbers: "1-4|"

SELECT "var1", "district_k", SUM("var2") AS "var3"
FROM (
  SELECT
    "var1",
    CASE
WHEN ("district_k" = '1301') THEN 'Arlesheim'
WHEN ("district_k" = '1302') THEN 'Laufen'
WHEN ("district_k" = '1303') THEN 'Liestal'
WHEN ("district_k" = '1304') THEN 'Sissach'
ELSE 'Waldenburg'
END AS "district_k",
    ""
  FROM (
````
````

Mit Resultierenden Daten Weiter Arbeiten

▼ Code

```
db_sql_small |>
  ggplot(
    aes(
      x = var1,
      y = var3,
      fill = fct_reorder(district_k, var3)
    )
  ) +
  geom_col() +
  labs(
    x = "",
    y = "Total",
    fill = "Bezirk"
  ) +
  scale_y_continuous(
    label = scales::label_number(prefix = "CHF ", big.mark = "'")
  ) +
  theme_minimal()
```

Resultierende Daten aus Datenbank Holen

```
1  class(df_db_small)
```

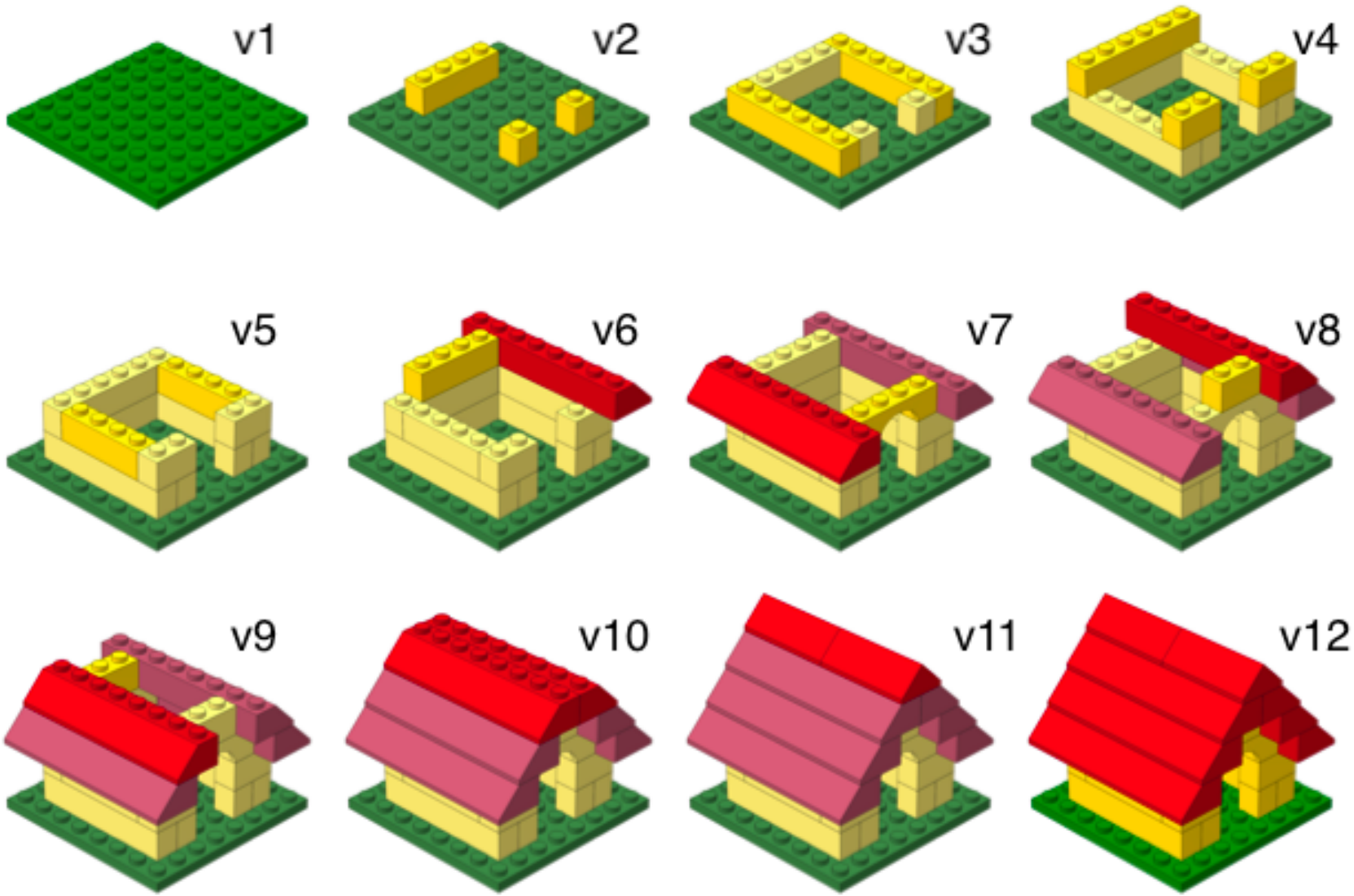
👉 Daten in DB (*lazy loading*)

```
1  df <- df_db_small |>
2    collect()
3
4  class(df)
```

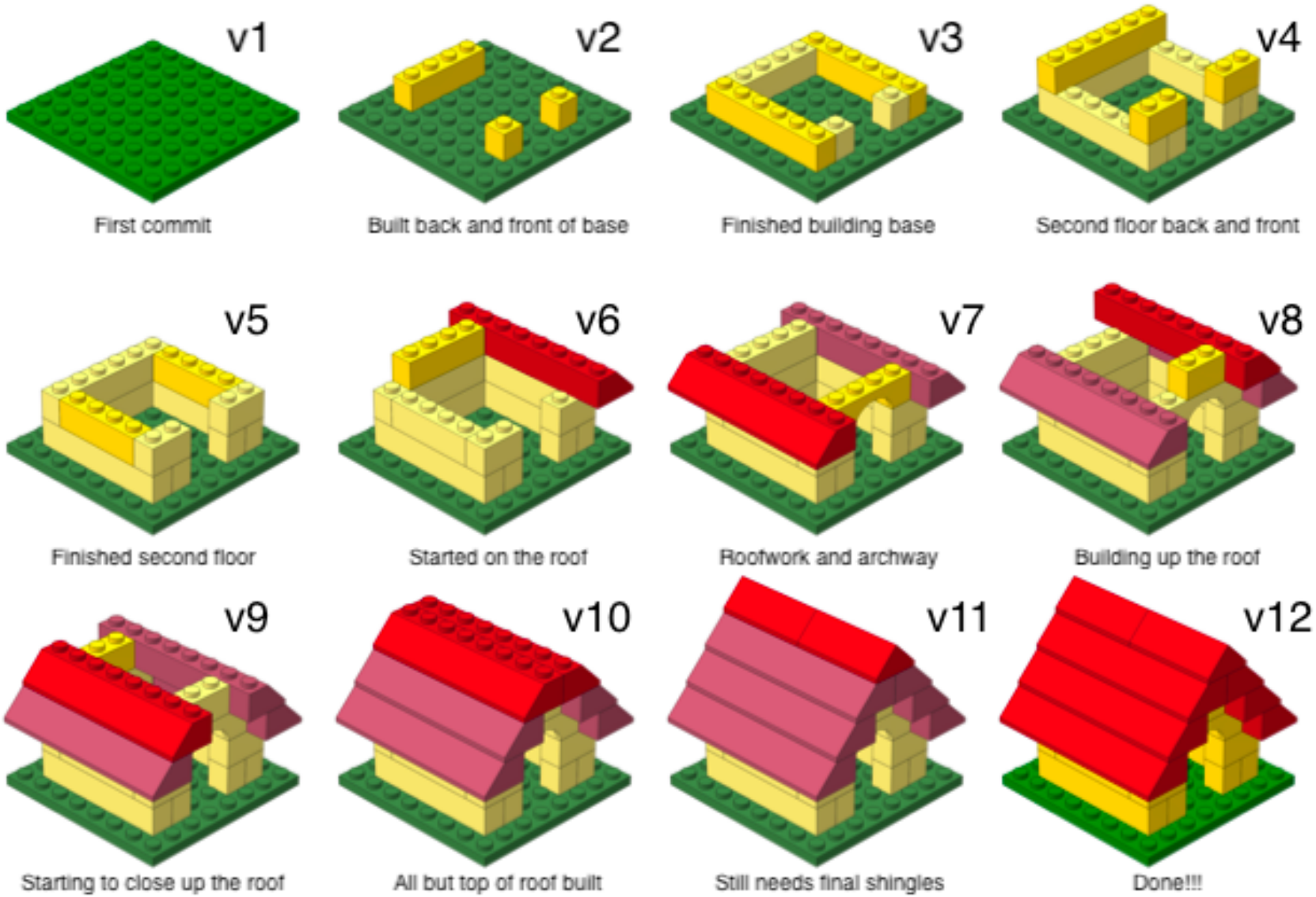
👉 Daten im Dataframe!

Versionierung und Kollaboration: `git`

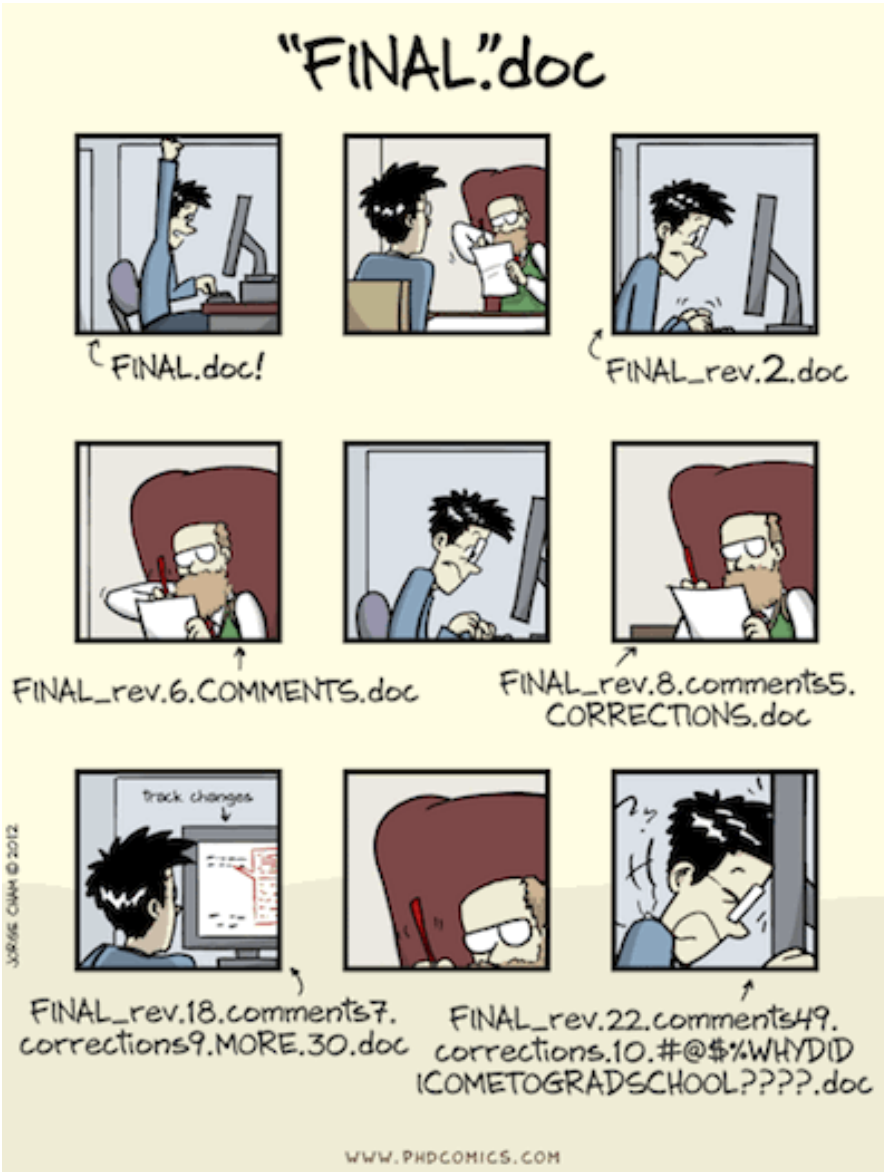
Versionierung



Versionierung

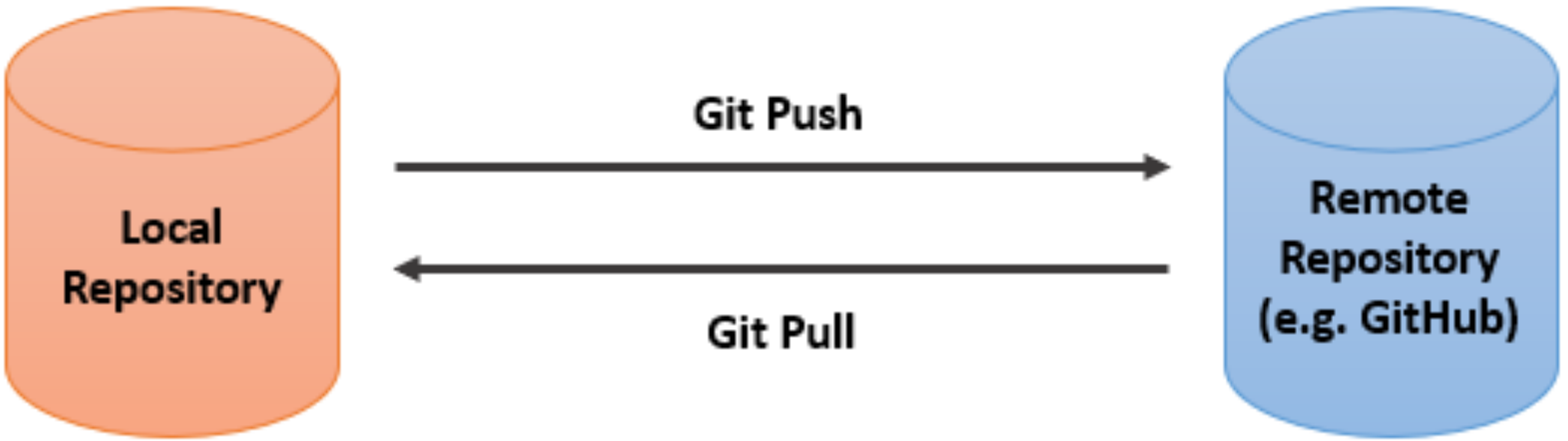


Warum Versionierung



Kollaboration mit sich Selbst

add, commit

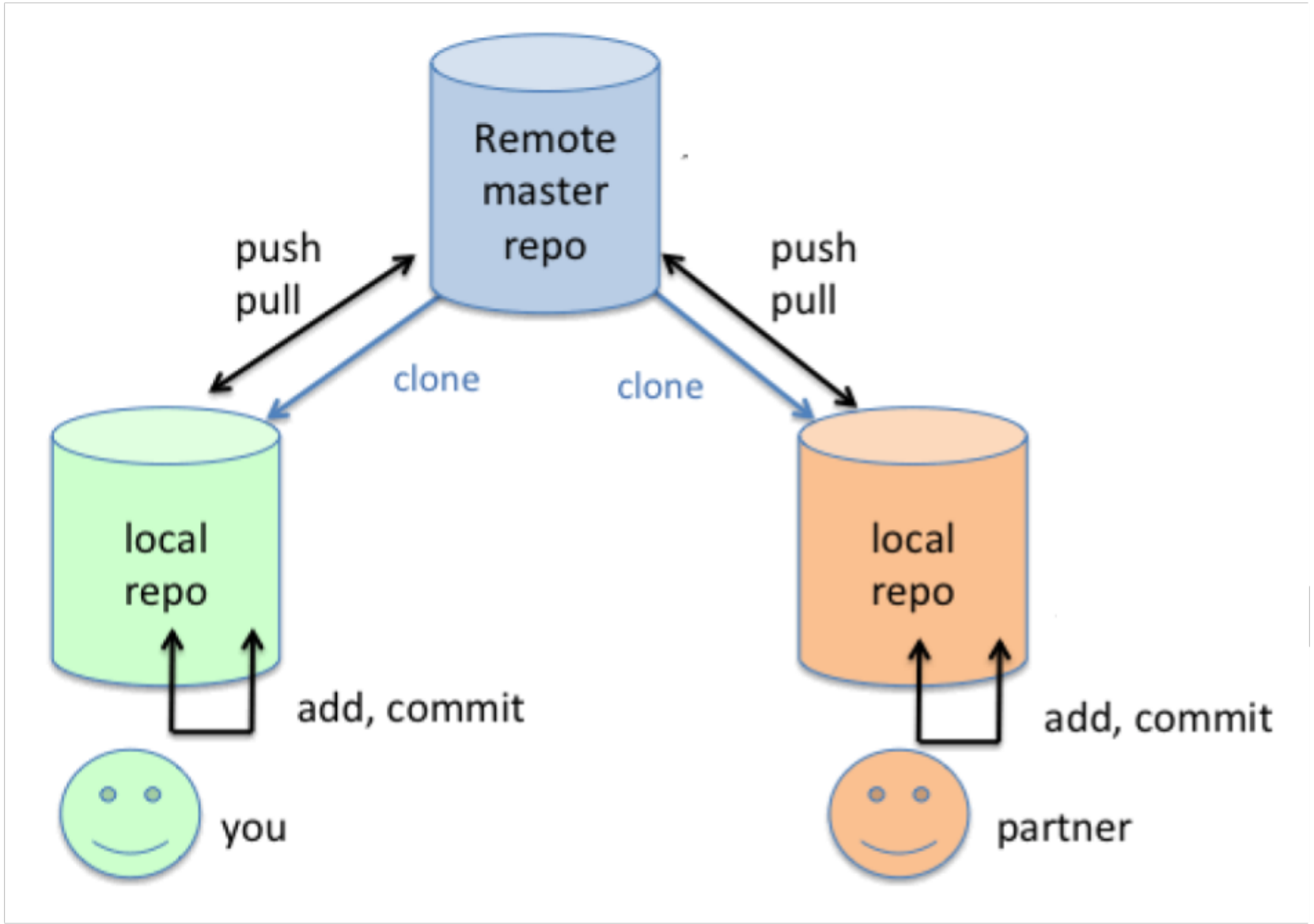


RStudio

github.com

BASEL LANDSCHAFT
FINANZ- UND KIRCHENDIREKTION
AMT FÜR DATEN UND STATISTIK

Kollaboration mit anderen

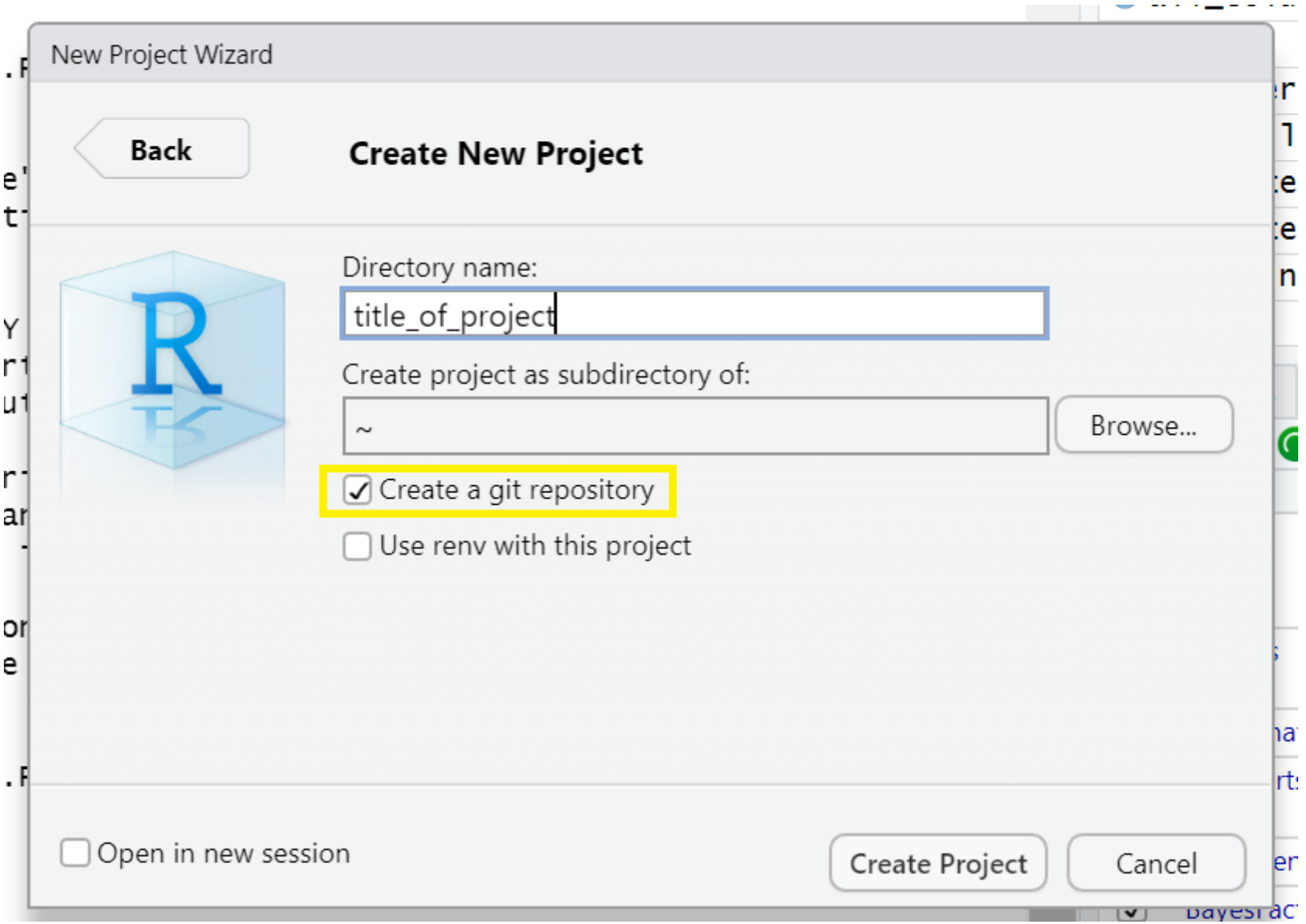


Warum Git und GitLab?

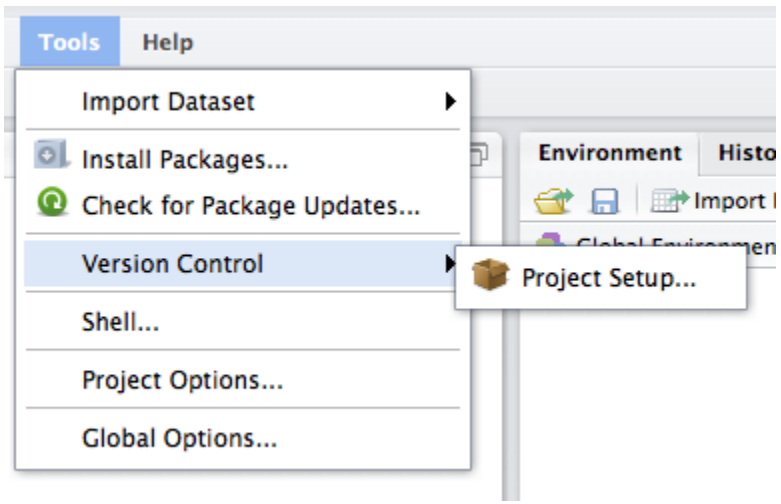
Git und GitLab

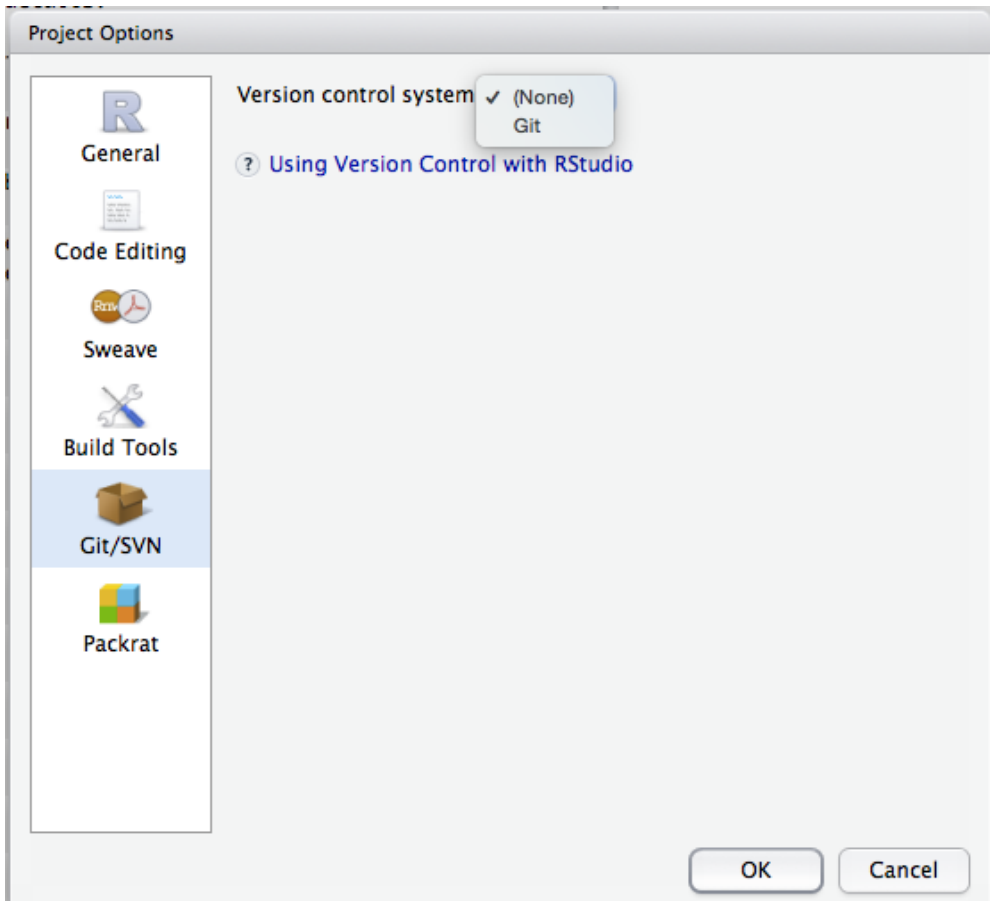
- (Nicht ganz) Millionen von Befehlen
- 99% der Zeit: `add`, `commit`, `push`, `pull`
- In RStudio oder *command line* (*terminal*)
- Happy Git and GitHub for the useR

Neues Projekt



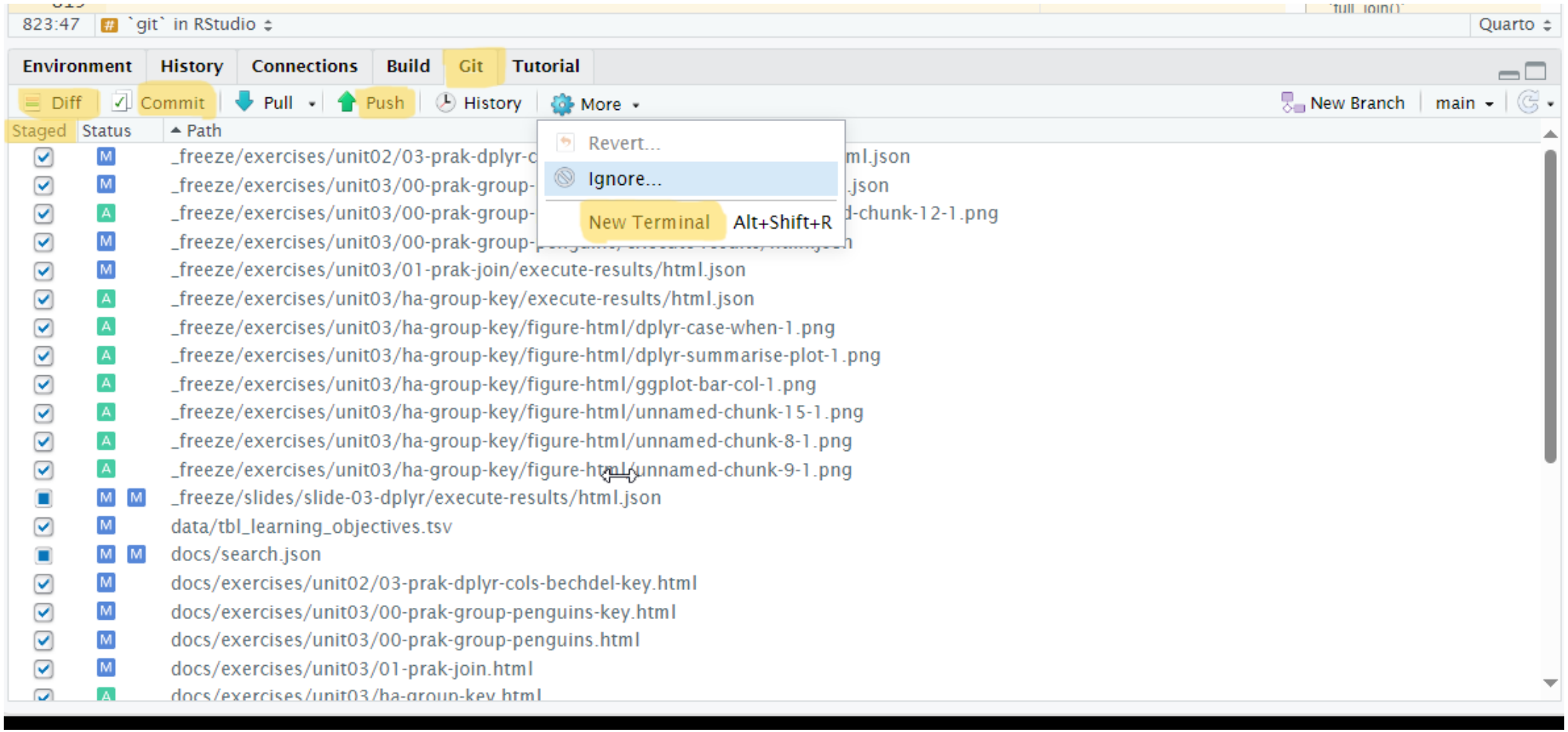
Existierendes Projekt





`git` in RStudio

Step-by-Step 1, step-by-step 2




THIS IS GIT. IT TRACKS COLLABORATIVE WORK ON PROJECTS THROUGH A BEAUTIFUL DISTRIBUTED GRAPH THEORY TREE MODEL.
COOL. HOW DO WE USE IT?
NO IDEA. JUST MEMORIZE THESE SHELL COMMANDS AND TYPE THEM TO SYNC UP. IF YOU GET ERRORS, SAVE YOUR WORK ELSEWHERE, DELETE THE PROJECT, AND DOWNLOAD A FRESH COPY.

Source: xkcd

Praktikum 06b: `git`

Quarto-Projekt mit `git`

20:00


at first I was like...
...but now it's like...
Horst '18

Fragen?

- R for Data Science
- DataCamp
- Stackoverflow
- ChatGPT, Perplexity Ai
- Souad Guemghar & Luca Hüsler

Danke! 

Slides created via revealjs and Quarto.

Access slides as PDF.

Git-Begriffe

- **repository** dein Projektordner
- **commit** ein Schnappschuss deines Repos
- **push** Commits an den *remote* senden
- **pull** Commits vom *remote* erhalten
- **clone** das Repository zum ersten Mal aus dem *remote* abrufen
- **remote** einen Computer, auf dem sich das Repository befindet