TALL
SOM FORTELLER

Introduction course

SUSIE JENTOFT, ASLAUG FOSS

DAY 1: 26 FEBRUAR, 2025

Statistisk sentralbyrå
Statistics Norway

Goals for course

- Gain an understanding for R and RStudio
- Open RStudio og run calculations
- Read in and manipulate data
- Create tables and summary statistics
- Create figures
- Edit and impute data
- Re-using code with loops
- Able to write own functions
- Able to create RMarkdown documents and visualize data in different ways
- Production organisation

Main themes

Wednesday 26.02	**Friday 28.02**	**Wednesday 05.03**	**Friday 05.03**	**Wednesday 19.03**	**Friday 21.03**
Introduction	Data manipulation	Data validation	Reusing code with loops	Documentation	Package management
Reading in data	Merging and plotting	Imputation	Writing functions	Github	Organising production code

Day structure

Time	Learning type
09:00-09:10	Review of previous day
09:10 – 09:45	Introduction to topic 1
09:45 – 10:15	Exercises
10:15 – 10:30	Coffee/tea break
10:30 – 11:15	Introduction to topic 2
11:15 – 11:50	Exercises
11:50 – 12:00	Summary

Susie Jentoft
(susie.jentoft@ssb.no)

Aslaug Foss
(aslaug.hurlen.foss@ssb.no)

Statistisk sentralbyrå
Statistics Norway

Course format and guidelines

- **Use chat** for asking questions or
- **Raise hand** to ask a question
- **Exercises:** Write or ask questions to us and eachother.
- Code for course is on **GitHub**
- **Mute microphone** when you are not speaking

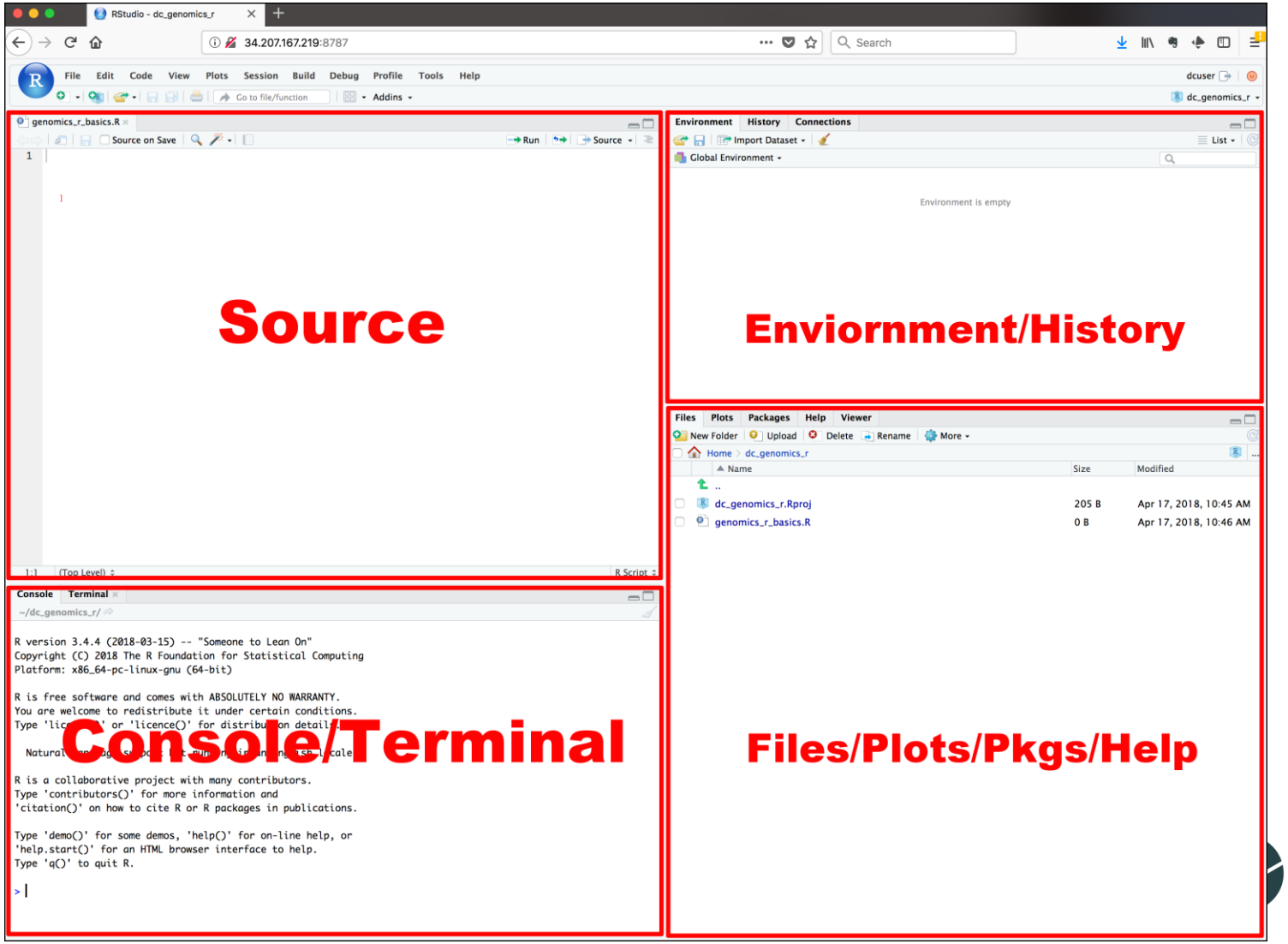
What is R ?

- Programming language and environment for statistics
- Developed by Ross Ihaka & Robert Gentleman (1993)
- Base + Packages
- RStudio:
  - IDE/development environment
  - Open source + commercial licenses

Why learn  ?

- Open source
- Large society and support
- Good graphics
- Good integration with other programmes
- Developed specifically for statistics
- Used within many statistical bureaus:
  - https://github.com/SNStatComp/awesome-official-statistics-software

RStudio


RStudio - dc_genomics_r
34.207.167.219:8787
Search
File Edit Code View Plots Session Build Debug Profile Tools Help
dcuser
Go to file/function
Addins
dc_genomics_r
genomics_r_basics.R
Source on Save
Run
Source
Source
Environment History Connections
Import Dataset
List
Global Environment
Environment is empty
Enviornment/History
Files Plots Packages Help Viewer
New Folder
Upload
Delete
Rename
More
Home dc_genomics_r
Name
Size
Modified
dc_genomics_r.Rproj
205 B
Apr 17, 2018, 10:45 AM
genomics_r_basics.R
0 B
Apr 17, 2018, 10:46 AM
Console Terminal
~/dc_genomics_r/
R version 3.4.4 (2018-03-15) -- "Someone to Lean On"
Copyright (C) 2018 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (64-bit)
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
Console/Terminal
Files/Plots/Pkgs/Help

Working with code files

Create a new file

- File > New File > R Script or 

Open an existing file

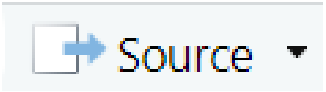
- File > Open File ... or 

Save file

- File > Save or 

Running code

- Write code in source/code files
- Run code by clicking on 

or ctrl + enter
  - This will run the line your mouse is on
  - Highlight an area to run a segment or several lines
  - Also possible to run whole code with 

- Lines that start with # are for comments (not run)
- Use 4x# to create headings (#### Heading ####)

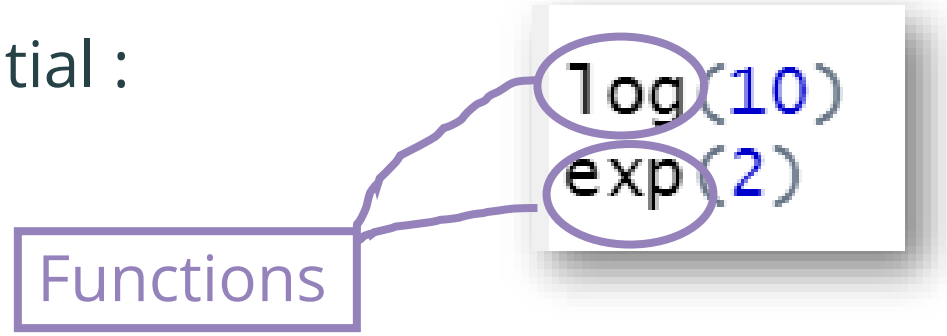
Basic operations

- Basic calculations

```
2 + 2
5 - 2
25 * 4
5 ^ 2
```

Base functions

Log and exponential :

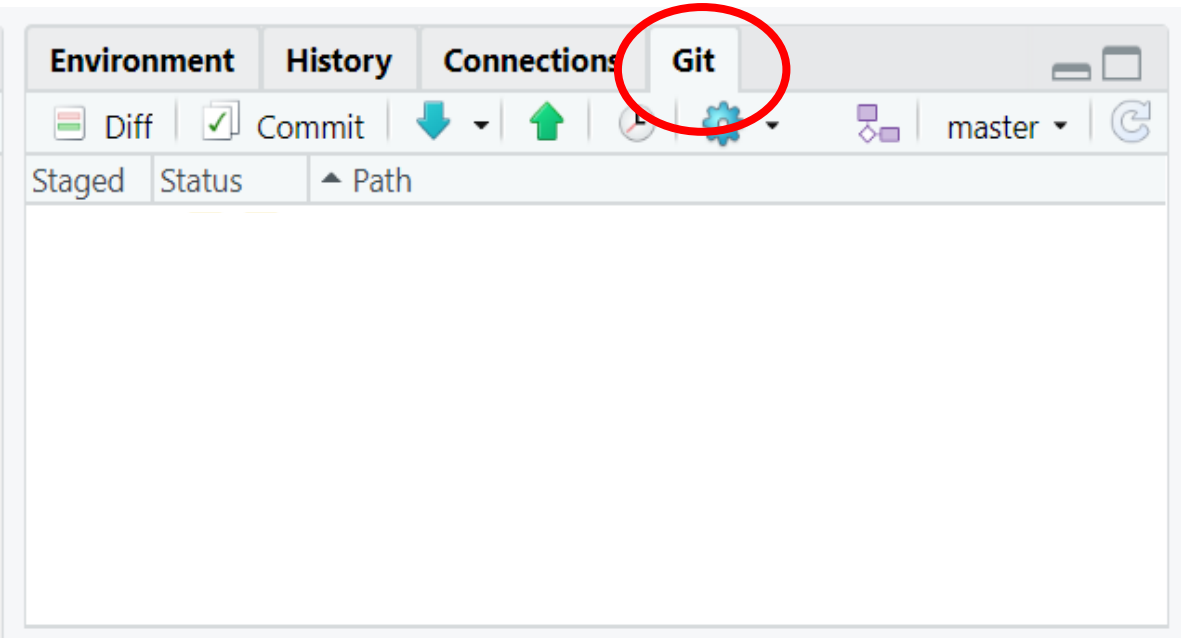


All functions in R have help files:

```
?log
```

Git and GitHub

- Git: version control for code
- Distribute code and manage projects with several contributors
- Git is well integrated in RStudio



Clone a repository

- **Github**: https://github.com/statisticsnorway/kurs-r-ukraine
  - Click on 

  - Copy HTTPS address (URL)
- In **RStudio**:
  - File > New project > Version Control > Git
  - Paste in address under «Repository URL»
  - Click Create Project
  - **Save files** you change (for example exercises) with a new name

Statistisk sentralbyrå
Statistics Norway

Example code for today

Example code for today is in the examples folder and is called:

Rcode_day1.R

Strings

```
"Hello, world!"
```

- Single (' ‘) or double quotes (" ") – be consistent!
- Combine two or more with `paste()`
- Take out part of a string:

```
substr(string, start, stopp)
```

- Print to console `print()`

Objects

- Creating an objects gives a name to a value/or string
- We use **<-** to allocate (give a name to) an object

```
x <- 2
```

- To see what is stored in the object, write and run the name

```
x
```

Objects

- CASE SENSITIVE
- Must start with a letter
- No spaces (use _)
- Different contents: number, string
- Can be written over/replaced

```
X  <- 5
```

```
Oblast ≠ oblast
```

Object type

- Find out type with

- Common object types:

Object type	Description
chr	Character/string
num	Number
Date	Date (use lubridate package)
Factor	Categorical variable (fixed levels)
data.frame	Dataset

Vectors

- Objects which hold several values
- Use c() to create vector (combine)
- Use comma to separate elements
- Must be same object type

```
c(1, 7, 10)

c("Kongsvinger", "Oslo")
```

Vectors as objects

- Give vectors a name with `<-`
- Find the length of a vector with `length()`
- Determine the object type with `str( )`

Access an element [ ]

- Use [ ] with an index number to fetch a value
- Indexing starts from 1
- Possible to index several values
- Exclude an value with “-”

```
vector_name[2]
vector_name[c(2,3)]
vector_name[2:4]
```

Exercise 1

- Clone repository for the course
- Open file Exercises_day1.R and do exercise 1

Logical statements/tests

- Compare objects
  - TRUE/FALSE

Code i R	Description
==	Compare if the same
!=	Compare if they are different
>	Greater than
>=	Greater than or equal
<	Less than
<=	Less than or equal
%in%	Is in/contains

- Can be used for single values or vectors
- Combine with **&** (and), **|** (or)

Create data

```
data.frame()
```

- Objects can also be data (with rows and columns)
- Combine different types of data (numbers and strings)

```
data.frame(object1, object2)
```

```
data.frame(column_name1 = c(1, 2), column_name2 = c("Industry", "Agriculture"))
```

Look at the data

- Write data name
- Click on dataset name under Environment
- Or

```
View()
```

Statistisk sentralbyrå
Statistics Norway

R Packages

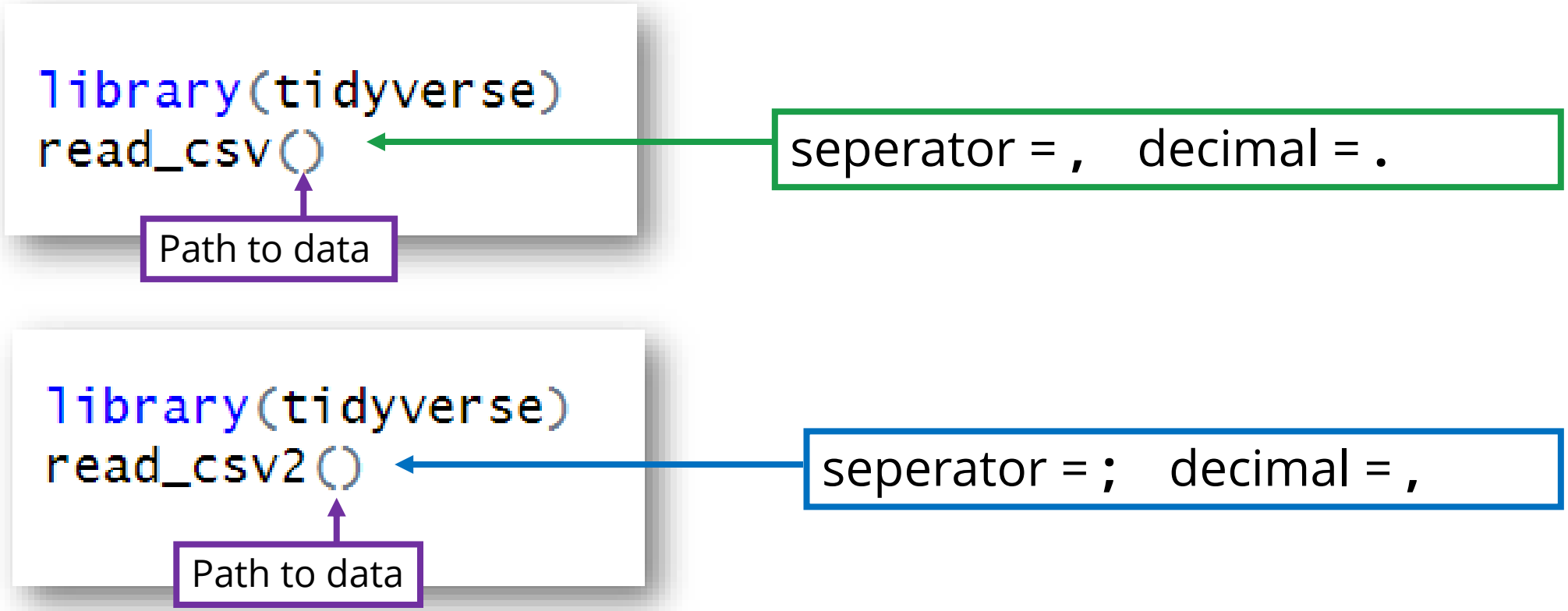
- Collection of code and functions
- CRAN (www.r-project.org)
- Install package one time

```
install.packages("package_name")
```

- Load package each time RStudio starts

```
library(package_name)
```

Read in data: .csv file



Read in data: stata file

- We can read in stata datsets direct to R using

```
library(haven)
read_dta()
```

Path to data

Read in dataset: fixed width format file

- Need to specify positions/widths. Use extra function: fwf_cols

```
library(tidyverse)
read_fwf(file, col_positions)
```

Path to data

Position or widths of variables

Look at the data

```
glimpse(dataset)
summary(dataset)
names(dataset)
```

Variables and type

Average, min, max, osv

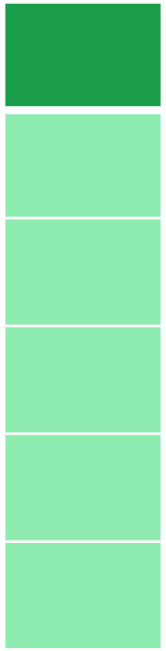
Variable names

Exercise 2:

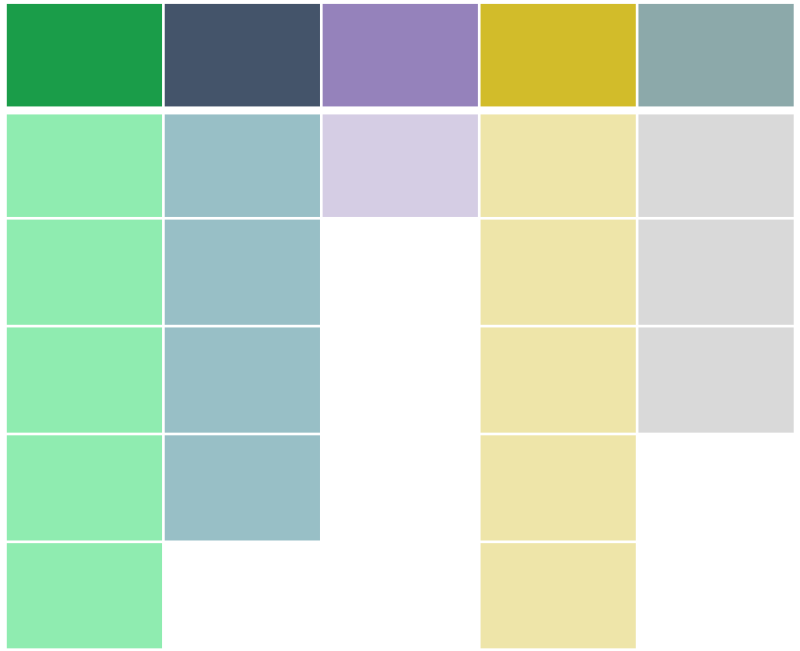
- Exercise 2 is in the file : **Exercises_day1.R**

Vectors, lists and data

Vector
`c()`



List
`list()`



Data
`Data.frame()`