

# Special Topics in Biostatistics and Bioinformatics Meet-and-greet

Ege Ülgen, M.D.  
egeulgen@gmail.com

3 March 2022



**ACIBADEM**  
MEHMET ALİ AYDINLAR  
ÜNİVERSİTESİ

# Ege Ülgen

- Email: [egeulgen@gmail.com](mailto:egeulgen@gmail.com)
- Office: Block A, 6<sup>th</sup> floor (Dept. of Biostatistics and Medical Informatics)
- Office hours: by appointment

# Lecture Notes and Scripts

- Available on GitHub:

[https://github.com/egeulgen/BB512 BB612 21 22](https://github.com/egeulgen/BB512_BB612_21_22)

# Syllabus

Week 1 – Introduction

Week 2 – Exploratory Data Analysis

Week 3 – Dimension Reduction/Preprocessing and Normalization

**Week 4 – Review of HW1**

Week 5 – Batch effects and Confounders

Week 6 – Inference

**Week 7 – Review of HW2**

# Syllabus (cont.)

Week 8 – Gene Set Enrichment

Week 9 – Gene Expression Analysis

**Week 10 – Review of HW3**

Week 11 – Next Generation Genomics Sequencing Analysis

**Week 14 – Review of Final Projects**

# Course Evaluation

- Homework – 30%
- Final Project – 70%

Software



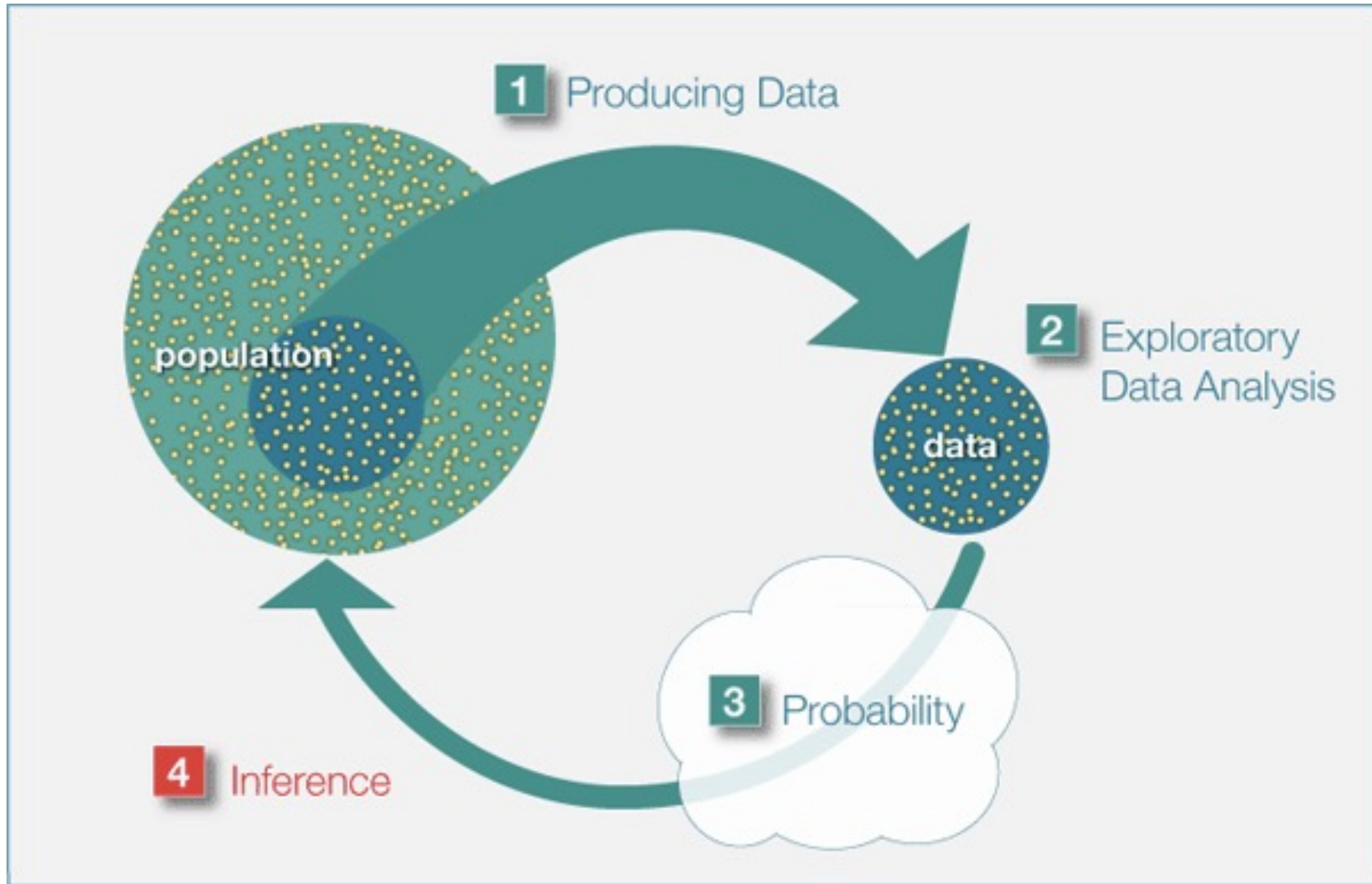
Studio<sup>®</sup>



Some Introduction

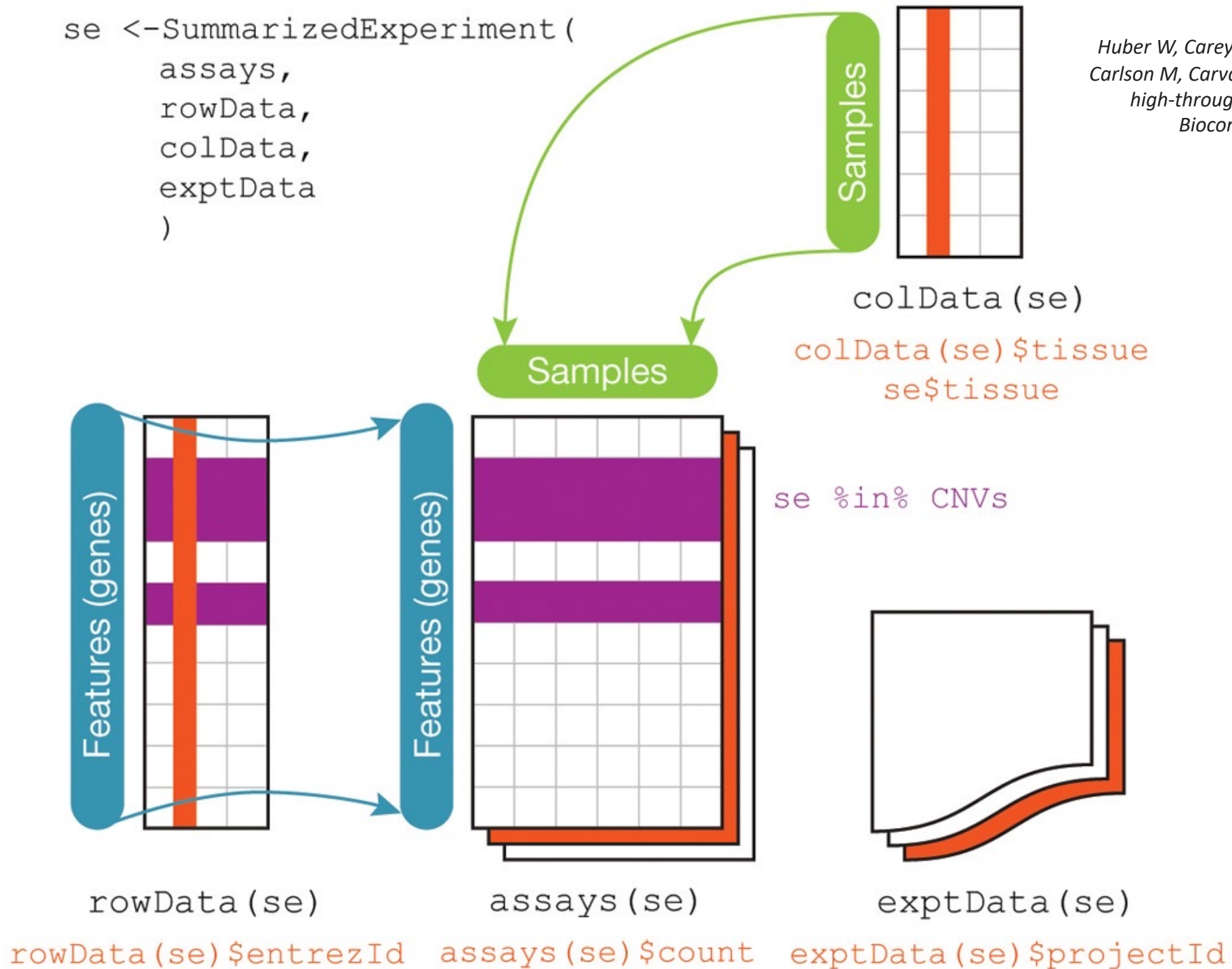


# Central Dogma of Statistics



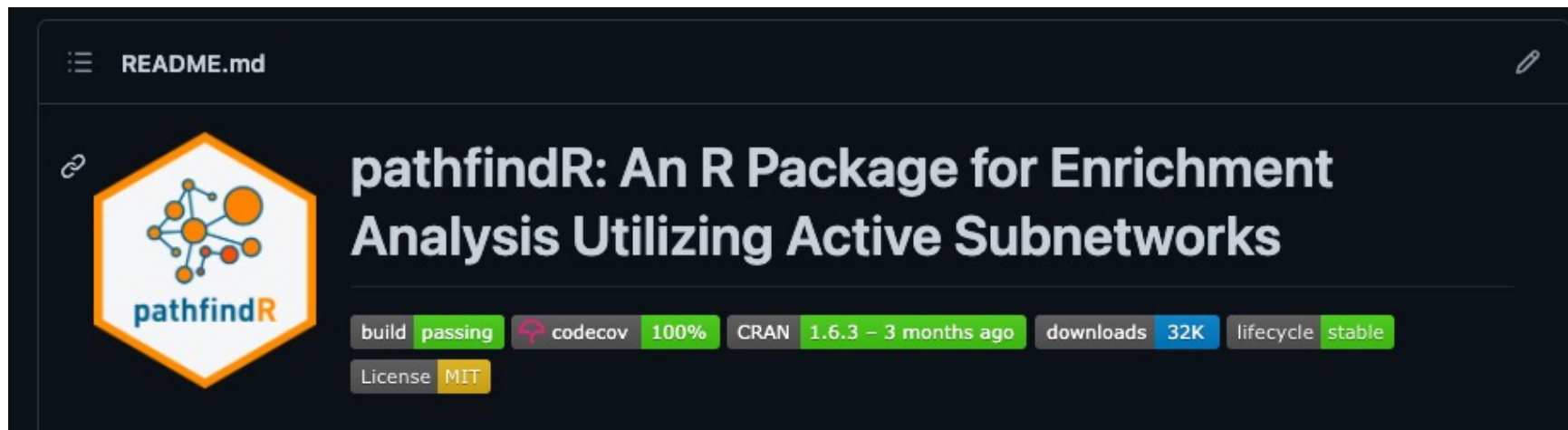
```
se <- SummarizedExperiment (
  assays,
  rowData,
  colData,
  exptData
)
```

Huber W, Carey VJ, Gentleman R, Anders S, Carlson M, Carvalho BS, et al. Orchestrating high-throughput genomic analysis with Bioconductor. Nat Methods. 2015 Feb;12(2):115–21.



# Where to find trustworthy methods

- Articles
- GitHub
- Forums
- ...



# Where to get help

- If you can't figure out how to solve an issue, Google is your friend.  
e.g., "how to calculate empirical p values"
- If you get an error and cannot fix it. C/P the error into Google.  
Someone else most likely had a similar problem
- Some good sources for asking about bioinformatics issues:
  - Biostars - <https://www.biostars.org/>
  - Bioconductor Support - <https://support.bioconductor.org/>
  - Stack overflow - <https://stackoverflow.com/>

# Reproducibility

- Clean data
- (A document with
  - variable names
  - descriptions
  - units ...)
- (preferably Rmd or similar) Code to analyze raw input data and generate the output
  - With clear descriptions/comments
- Organized project directory

# «Clean» Data

country	year	cases	population
Afghanistan	1999	7745	19987071
Afghanistan	2000	8666	20595360
Brazil	1999	37737	172006362
Brazil	2000	80488	174504898
China	1999	212258	1272915272
China	2000	213766	1280425583

variables

country	year	cases	population
Afghanistan	1999	7745	19987071
Afghanistan	2000	8666	20595360
Brazil	1999	37737	172006362
Brazil	2000	80488	174504898
China	1999	212258	1272915272
China	2000	213766	1280425583

observations

country	year	cases	population
Afghanistan	99	7745	19987071
Afghanistan	00	8666	20595360
Brazil	99	37737	172006362
Brazil	00	80488	174504898
China	99	212258	1272915272
China	00	213766	1280425583

values