

# CBB Workshop

How to create a “table 1” and produce publication-ready figures in R

Thursday 23 May 2024, 13:00 – 16:00

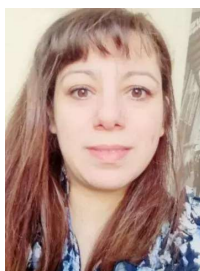
# KI Center for Bioinformatics and Biostatistics (CBB)



**Carsten Daub**  
director



**Hannes Hagström**  
co-director



**Ghada Nouairia**  
bioinformatician



**Tassos Damdimopoulos**  
bioinformatician

**CBB drop-in:**

Thursdays 13:00 – 15:00

NEO Protein

**cbb@ki.se**



**Henrike Häbel**  
biostatistician



**Rickard Strandberg**  
biostatistician



**Davide Valentini**  
biostatistician



**Michael Ingre**  
biostatistician



**Laura Vossen Engblom**  
biostatistician

# Agenda

13:00 – 13:15	Introduction
13:15 – 14:00	Part 1: how to create a table 1 (baseline characteristics)
14:00 – 14:30	Coffee break
14:30 – 16:00	Part 2: data visualization using ggplot2

## Course materials

- <https://github.com/cbb-ki/Workshops>
  - Save to your **local** harddisk
- Open the script "CBB\_table1\_ggplot.R"
- On line 37, change the directory
- Run line 1 – 40:
  - Install R packages
  - Load R packages
  - Read in the data

# Example

**TABLE 1. Demographics and Baseline Characteristics**

		Cilofexor 100 mg (n = 22)	Cilofexor 30 mg (n = 20)	Placebo (n = 10)	Total (n = 52)
Demographics	Age (years)	43 (36, 47)	46 (35, 57)	39 (33, 52)	43 (35, 52)
	Male, n (%)	11 (50%)	14 (70%)	5 (50%)	30 (58%)
	White, n (%)	17 (77%)	15 (75%)	7 (70%)	39 (75%)
	Diabetes, n (%)	6 (27%)	2 (10%)	1 (10%)	9 (17%)
	Weight (kg)	73.5 (67.5, 89.1)	79.8 (68.4, 95.9)	82.2 (63.0, 83.3)	77.9 (67.4, 88.6)
	BMI (kg/m <sup>2</sup> )	25.8 (23.2, 30.3)	25.9 (22.8, 29.9)	25.8 (23.9, 29.6)	25.8 (23.2, 29.9)
	IBD, n (%)	13 (59%)	11 (55%)	7 (70%)	31 (60%)
	UDCA, n (%)	10 (46%)	9 (45%)	5 (50%)	24 (46%)
Liver biochemistry	ALP (U/L)	350 (312, 387)	344 (271, 460)	380 (265, 547)	348 (288, 439)
	ALP (× ULN)	2.87 (2.45, 3.51)	2.73 (2.15, 3.74)	3.31 (2.05, 4.33)	2.78 (2.29, 3.66)
	GGT (U/L)	305 (192, 542)	564 (255, 910)	377 (224, 622)	423 (203, 628)
	Total bilirubin (mg/dL)	0.6 (0.5, 1.1)	0.8 (0.6, 1.0)	0.6 (0.5, 0.9)	0.7 (0.5, 1.0)
	ALT (U/L)	110 (83, 156)	119 (60, 197)	77 (59, 123)	109 (63, 156)
	ALT (× ULN)	2.90 (1.92, 4.46)	3.01 (1.74, 4.81)	2.01 (1.60, 2.86)	2.67 (1.70, 4.12)
	AST (U/L)	67 (52, 98)	75 (44, 104)	59 (47, 76)	64 (47, 99)
	AST (× ULN)	1.92 (1.49, 2.87)	2.11 (1.29, 2.90)	1.74 (1.31, 2.10)	1.83 (1.34, 2.78)
	Albumin (g/dL)	4.4 (4.2, 4.5)	4.5 (4.2, 4.7)	4.6 (4.2, 4.7)	4.4 (4.2, 4.7)
Fibrosis and inflammation	ELF	9.26 (8.73, 9.66)	9.77 (9.26, 10.31)	9.09 (8.87, 9.60)	9.38 (8.91, 9.88)
	FibroTest	0.29 (0.27, 0.44)	0.47 (0.30, 0.57)	0.34 (0.23, 0.51)	0.40 (0.28, 0.51)

Trauner et al. 2019  
Hepatology 70(3)

## The caffeine dataset

- In the experiment, 100 participants were given either green tea or coffee, in doses of either 1 or 5 servings. They performed a cognitive test where higher scores indicate better performance.

*Dataset from [bookdown.org](https://bookdown.org)*



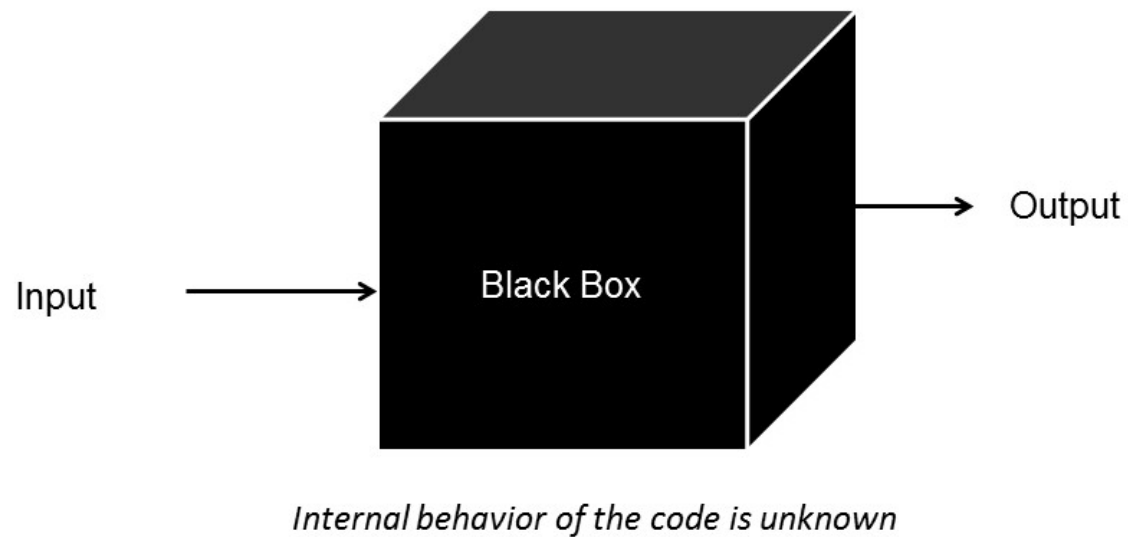
# Let's get started...

## Take home messages

- Free HarvardX course ["Data science: R basics"](#)
- Package websites: cheat sheets, vignettes, tutorials
- R command to create a table 1: `tbl_summary(my.data)`
- ggplot2 breaks up graphs into semantic components
  - Same logic in most R visualization packages



## Can chatGPT do this?



# Thank you for attending!

## Why learn R?

- Free
  - Latest statistical and bioinformatical methods
  - All-in-one
  - Reproducible
  - Better statistical understanding
- 
- Free HarvardX course ["Data science: R basics"](#)