

Introductie programmeren met Rstudio

Bram van Dijk (hij/hem)



Bespreking werkcollege

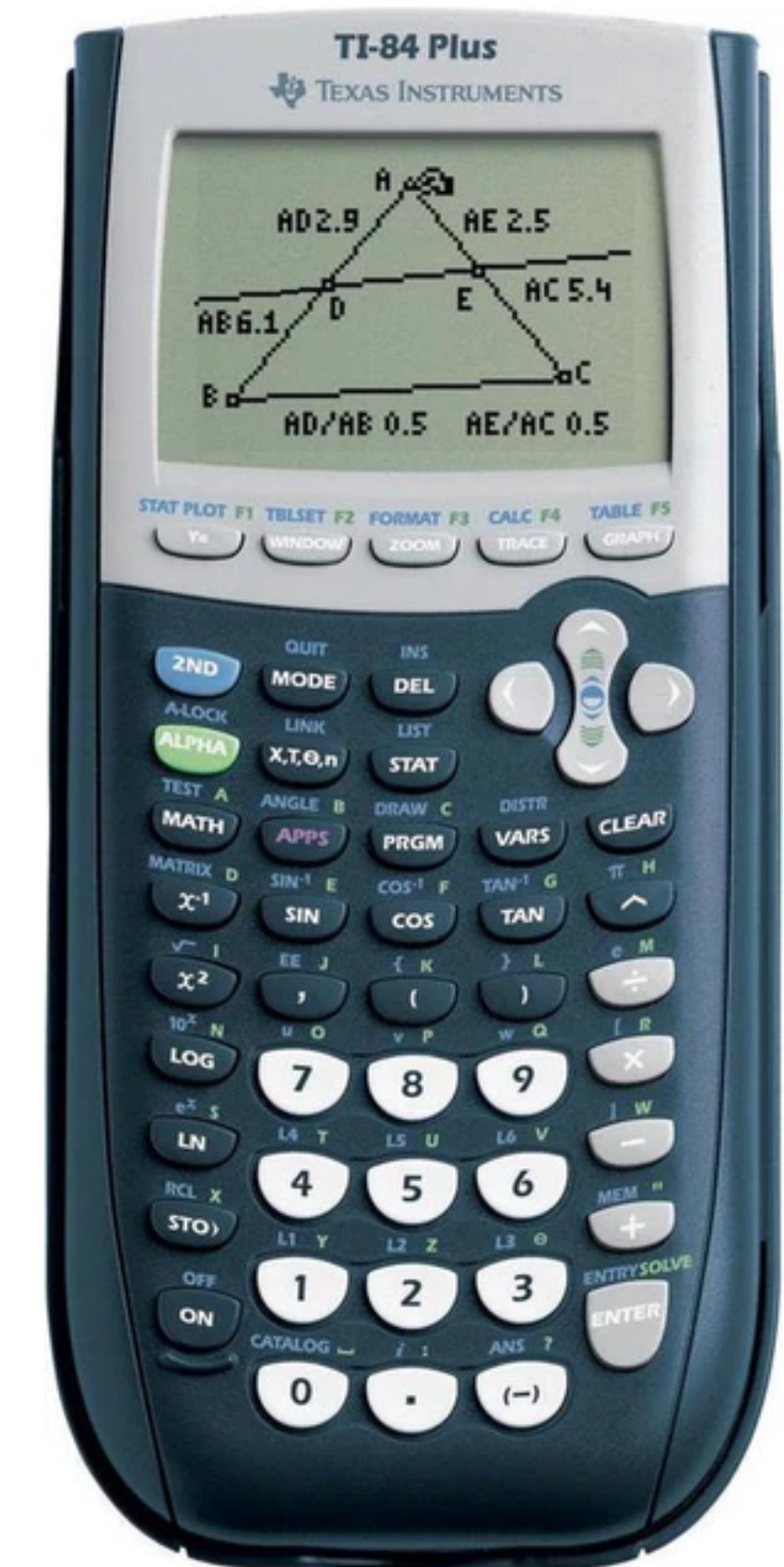
Oefening 1.1 (Rekenen met R)



Maak een nieuw (leeg) Rscript in Rstudio, en voer de volgende berekeningen uit:

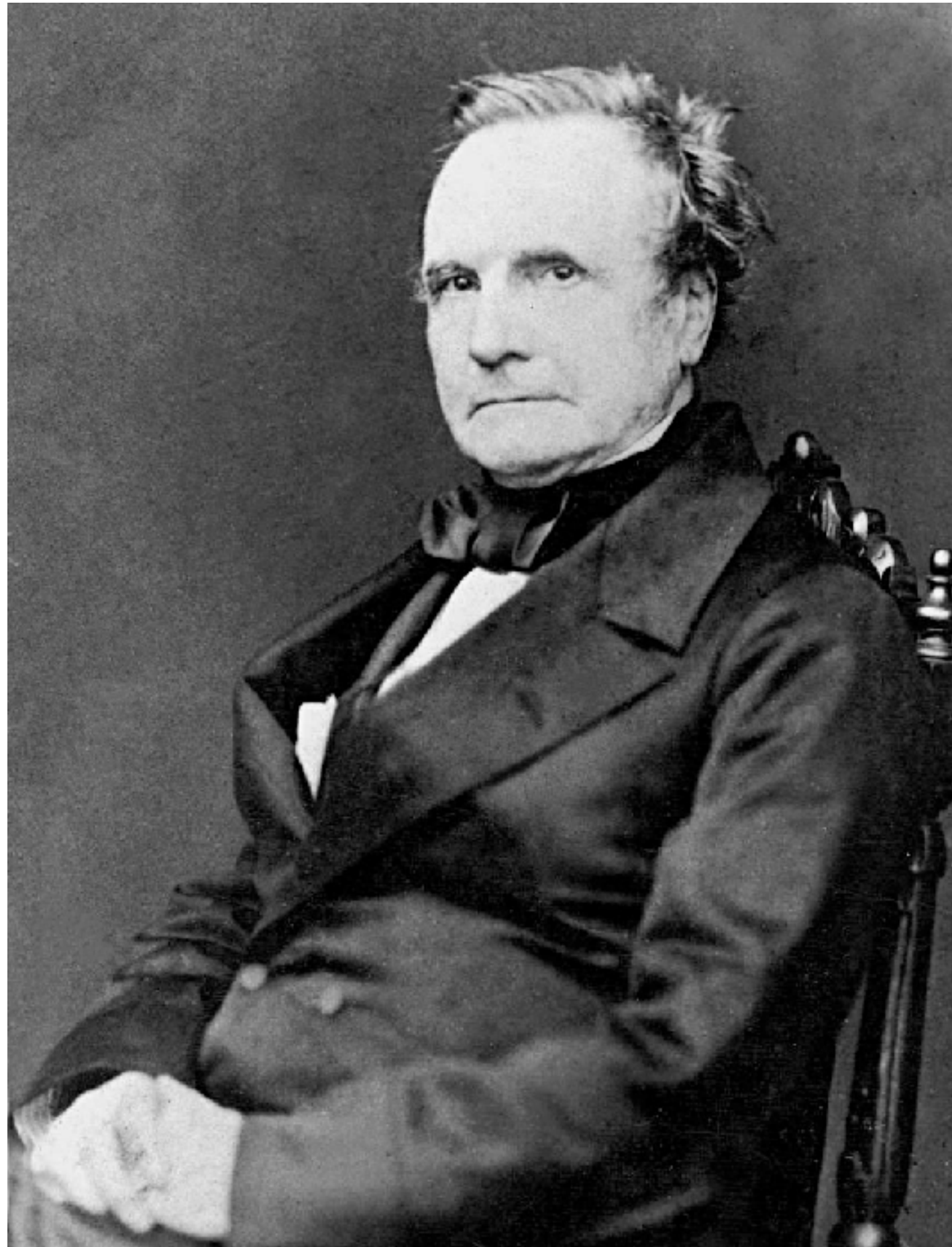
- Definieer $a=2$, en reken uit wat $a \cdot a \cdot a = a^3$ is.
- Definieer nu $b=9.5$ en reken uit wat b^{19} is (let op: je hoeft niet 19 keer b te typen)
- Stel je zet eerst 3 stappen naar links (laten we deze afstand a noemen), dan 4 stappen naar voren (laten we deze b noemen), en loopt vervolgens in een rechte lijn terug naar je beginpositie. Schrijf een script met R die uitrekent hoe ver je in totaal hebt gelopen. Zorg dat de uitkomst van je berekening **niet** in de environment staat, maar slechts in de console wordt weergegeven.
- Bereken nu uit hoever je gelopen zou hebben als je in plaats hiervan 5 stappen naar links en 6 stappen naar boven had gezet. Zorg dit keer dat de uitkomst **wel** in de environment staat onder de naam "afstand". Voeg de onderstaande regel code toe aan het eind van je script (hint: er is een knop om de code te kopiëren!). Voer je script uit. Hoeveel stappen heb je gelopen? Sla je script op onder de naam **Pythagoras.R**.

```
print(paste('Je hebt', afstand, 'stappen gelopen'))
```



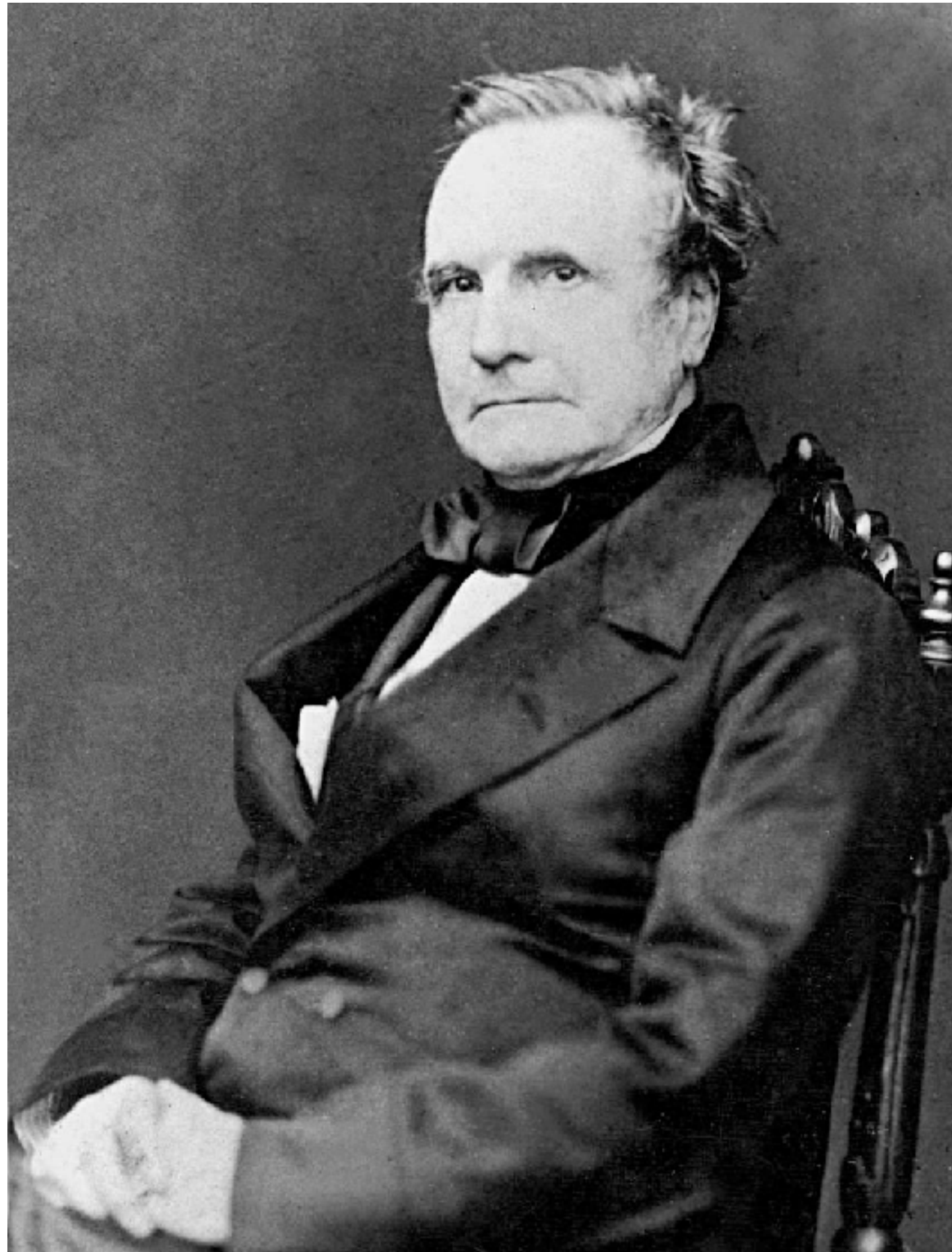
Maar R is meer dan een rekenmachine

Maar R is meer dan een rekenmachine



Charles Babbage (1791 – 1871)

Maar R is meer dan een rekenmachine

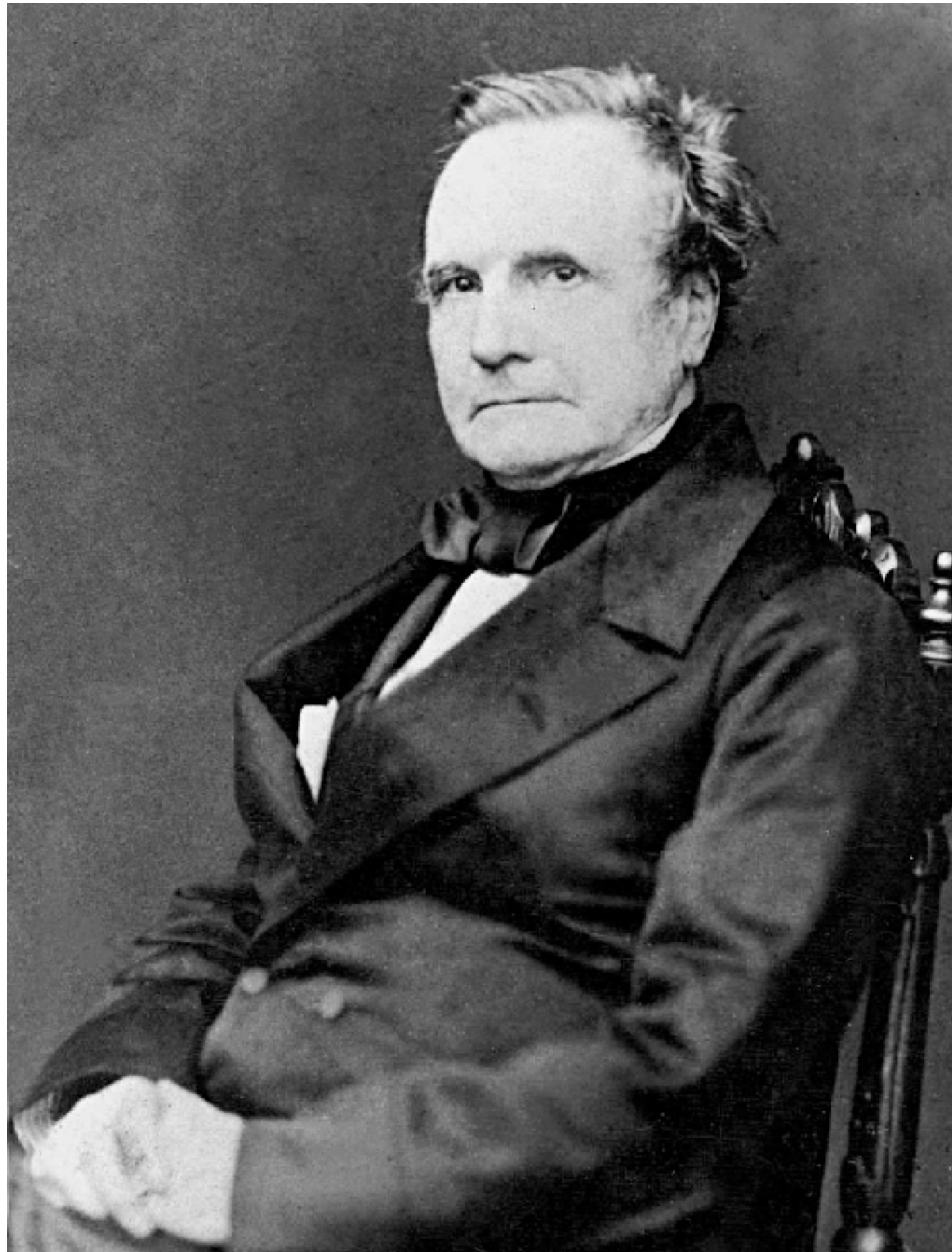


Charles Babbage (1791 – 1871)



**Benadryl Cabbagepatch
as Alan Turing**

Maar R is meer dan een rekenmachine



Charles Babbage (1791 – 1871)



**Benadryl Cabbagepatch
as Alan Turing**



Alan Turing (1912 – 1954)

Maar R is meer dan een rekenmachine



Alan Turing (1912 – 1954)

Maar R is meer dan een rekenmachine

A computer can be said to “think” if a human interrogator could not tell it apart (through conversation) from a human being.



Alan Turing (1912 – 1954)

Maar R is meer dan een rekenmachine

A computer can be said to “think” if a human interrogator could not tell it apart (through conversation) from a human being.

“Machines take me by surprise with great frequency.”



Alan Turing (1912 – 1954)

Variabelen



Data structuren in R

vector
(*numeric*)

1

3

2

6

vector
(*string*)

a

c

bla

a

matrix
(*numeric*)

1

3

7

8

2

0

3

6

2

4

2

4

6

7

7

3

dataframe
(*columns are vectors*)

1

a

T

8.7

4

c

F

0.6

2

bla

T

4.0

6

a

T

1.4

Data structuren in R

- Datastructuren zijn alle manieren om verzamelingen van variabelen op te slaan. In R kan dit vooral met **vectoren**, **matrices**, en **dataframes**.

vector
(*numeric*)

1

3

2

6

vector
(*string*)

a

c

bla

a

matrix
(*numeric*)

1

3

7

8

2

0

3

6

2

4

2

4

6

7

7

3

dataframe
(*columns are vectors*)

1

a

T

8.7

4

c

F

0.6

2

bla

T

4.0

6

a

T

1.4

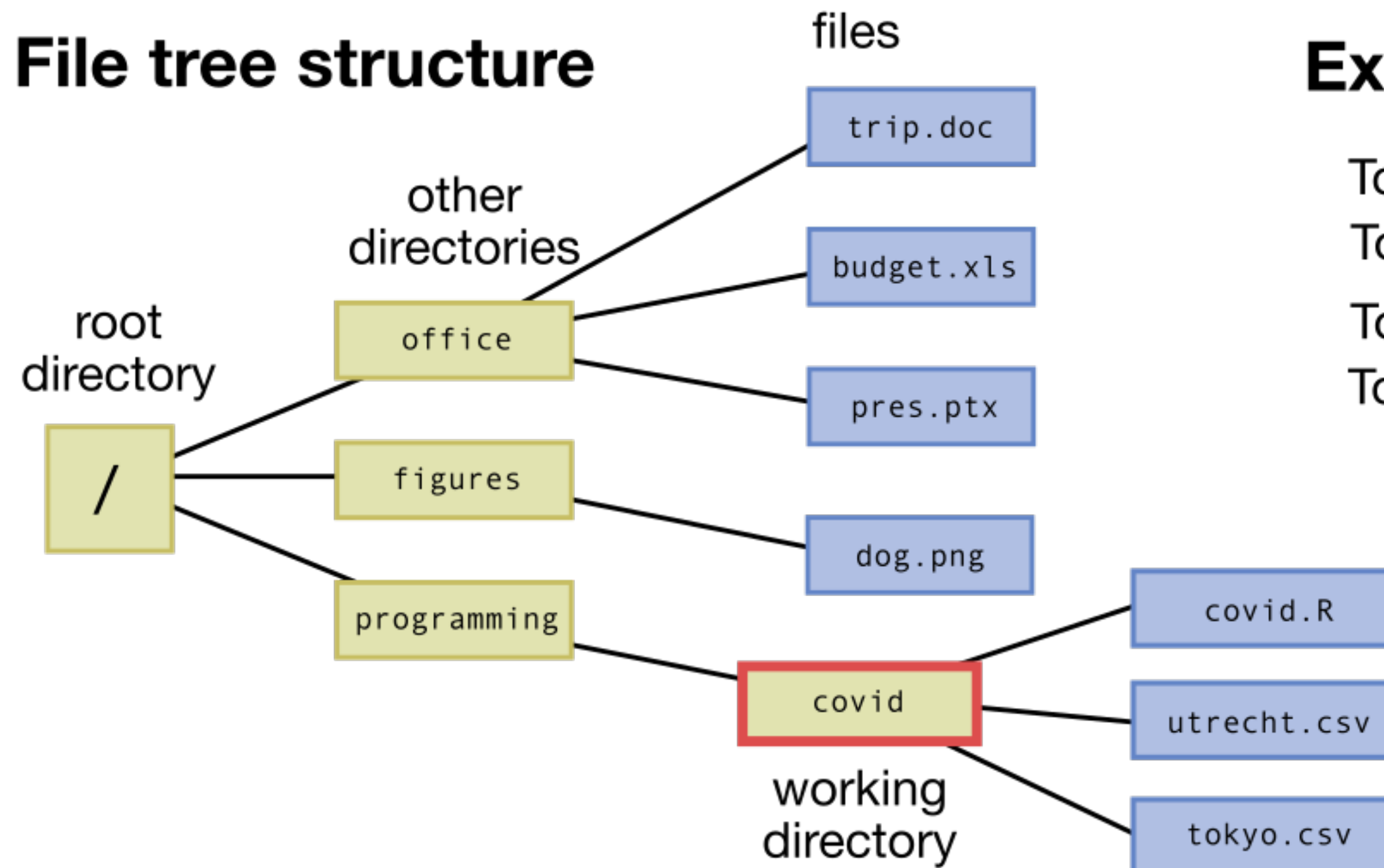
Comma-separated values bestand

covid_utrecht.csv

```
1] datapunt, week, maand, jaar, positieve_testen
2] 1, 1, 10, 2021, 221
3] 2, 2, 10, 2021, 271
4] 3, 3, 10, 2021, 344
5] 4, 4, 10, 2021, 507
6] 5, 5, 10, 2021, 514
7] 6, 1, 11, 2021, 837
8] 7, 2, 11, 2021, 1254
...
```

Files, paden, en 'working directory'

File tree structure



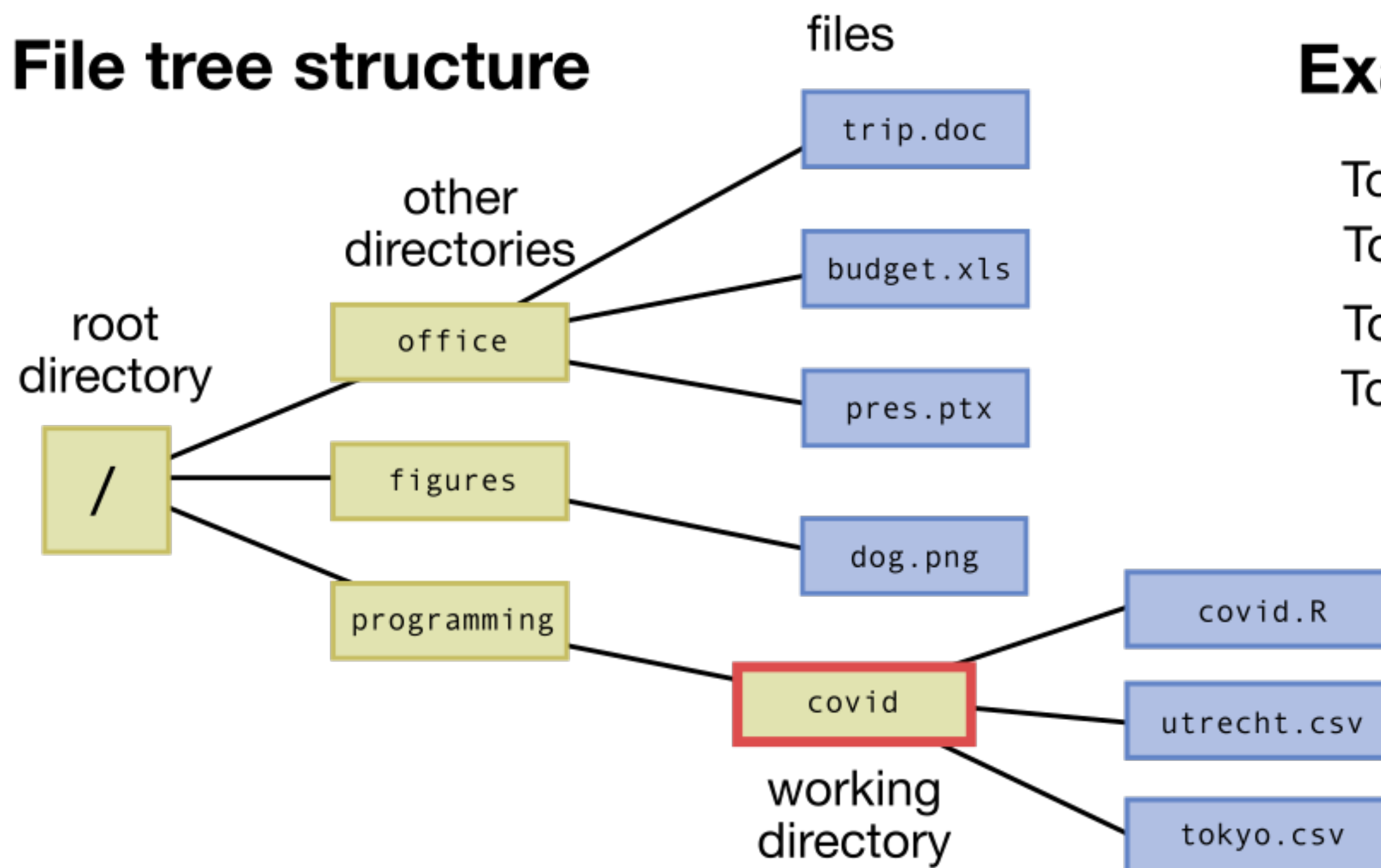
Example paths

To utrecht.csv (absolute): `/data/covid/utrechts.csv`
To utrecht.csv (relative): `utrecht.csv`
To dog.png (absolute): `/figures/dog.png`
To dog.png (relative): `../../../../figures/dog.png`

Files, paden, en 'working directory'

- Bestanden inlezen doe je meestal niet vanaf het internet, maar vanaf een mapje op je eigen computer. R moet dan wel weten waar de file staat.

File tree structure



Example paths

To utrecht.csv (absolute):	<code>/data/covid/utrechts.csv</code>
To utrecht.csv (relative):	<code>utrecht.csv</code>
To dog.png (absolute):	<code>/figures/dog.png</code>
To dog.png (relative):	<code>../../../../figures/dog.png</code>



Probleemoplossend denken en ‘pseudocode’

Seq 1: “ATAAGTACTGAGCGGTATGGTCGTAGTCGTGATGTCGTA”

Seq 2: “ATAAGGACTGAGCGGTATGGTCGTAGTCGTGTCATGTTGTA”

Pos : 123456789...



Probleemoplossend denken en ‘pseudocode’

Seq 1: “ATAAGTACTGAGCGGTATGGTCGTAGTCGTGATGTCGTA”

Seq 2: “ATAAGGACTGAGCGGTATGGTCGTAGTCGTGTCATGTTGTA”

Pos : 123456789...



Probleemoplossend denken en ‘pseudocode’

Seq 1: “ATAAGTACTGAGCGGTATGGTCGTAGTCGTGATGTCGTA”

Seq 2: “ATAAGGACTGAGCGGTATGGTCGTAGTCGTGTCATGTTGTA”

Pos : 123456789...



PSEUDOCODE:

we beginnen met 0 mismatches

voor elke positie van 1 t/m de **lengte van de sequentie**:

hetzelfde? doe niets

verschillend? tel 1 bij aantal mismatches op

klaar? geef het aantal mismatches weer

Probleemoplossend denken en ‘pseudocode’

Seq 1: “ATAAGTACTGAGCGGTATGGTCGTAGTCGTGATGTCGTA”
Seq 2: “ATAAGGACTGAGCGGTATGGTCGTAGTCGT CATGTTGTA”
Pos : 123456789...



PSEUDOCODE:

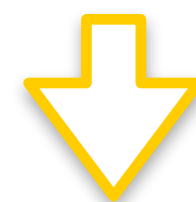
we beginnen met 0 mismatches

voor elke positie van 1 t/m de **lengte van de sequentie**:

hetzelfde? doe niets

verschillend? tel 1 bij aantal mismatches op

klaar? geef het aantal mismatches weer



- Hoe bepaal ik in R de lengte van een sequentie?
- Hoe kijk ik naar de waarde van de i-de positie?

Probleemoplossend denken en ‘pseudocode’

Seq 1: “ATAAGTACTGAGCGGTATGGTCGTAGTCGTGATGTCGTA”
Seq 2: “ATAAGGACTGAGCGGTATGGTCGTAGTCGT CATGTTGTA”
Pos : 123456789...

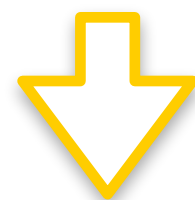


PSEUDOCODE:

we beginnen met 0 mismatches

voor elke positie van 1 t/m de **lengte van de sequentie**:
hetzelfde? doe niets
verschillend? tel 1 bij aantal mismatches op

klaar? geef het aantal mismatches weer



PSEUDOCODE 2:

zet de ‘strings’ om in ‘vectoren’ van karakters

gebruik != om te kijken waar de vectoren niet matchen

sommeer de uitkomst (want TRUE is 1!)

- Hoe bepaal ik in R de lengte van een sequentie?
- Hoe kijk ik naar de waarde van de i-de positie?

Probleemoplossend denken en ‘pseudocode’

Seq 1: “ATAAGTACTGAGCGGTATGGTCGTAGTCGTGATGTCGTA”
Seq 2: “ATAAGGACTGAGCGGTATGGTCGTAGTCGT CATGTTGTA”
Pos : 123456789...

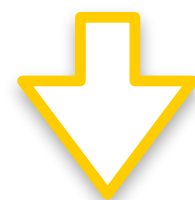


PSEUDOCODE:

we beginnen met 0 mismatches

voor elke positie van 1 t/m de **lengte van de sequentie**:
hetzelfde? doe niets
verschillend? tel 1 bij aantal mismatches op

klaar? geef het aantal mismatches weer



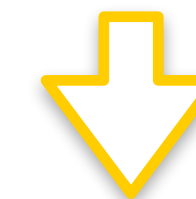
- Hoe bepaal ik in R de lengte van een sequentie?
- Hoe kijk ik naar de waarde van de i-de positie?

PSEUDOCODE 2:

zet de ‘strings’ om in ‘vectoren’ van karakters

gebruik != om te kijken waar de vectoren niet matchen

sommeer de uitkomst (want TRUE is 1!)



- Hoe zet ik een ‘string’ om in een ‘vector’?

Programmeren en AI

- Heel handig, maar vooral als je al een beetje weet hoe je moet programmeren
- Het kan ook R-code analyseren! (**niet** uitvoeren)
- Weet dat AI ook **veel energie kost**.



ChatGPT

Programmeren en AI

- Heel handig, maar vooral als je al een beetje weet hoe je moet programmeren
- Het kan ook R-code analyseren! (**niet** uitvoeren)
- Weet dat AI ook **veel energie kost**.



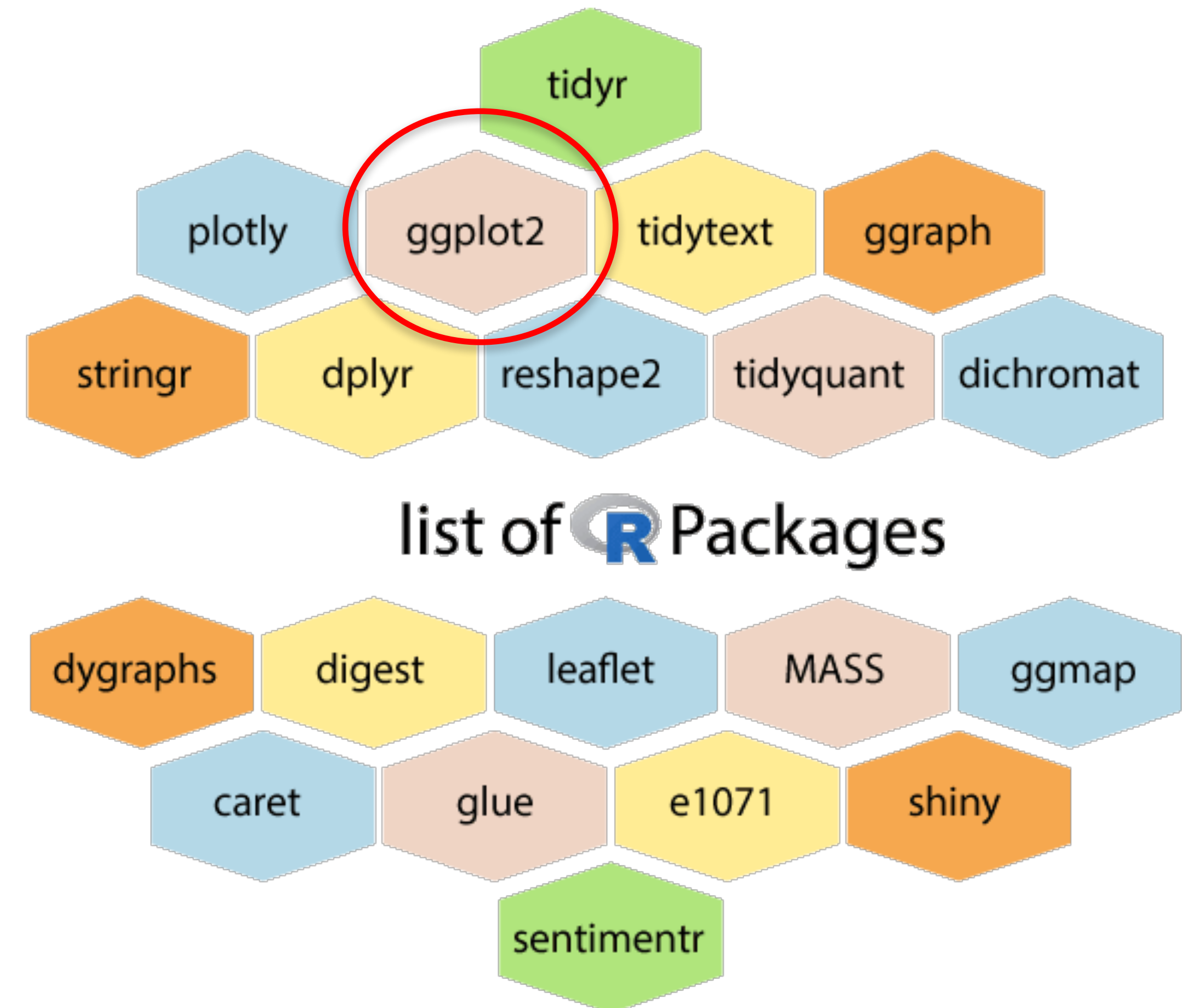
ChatGPT

The pace of the global energy transition has been vastly underestimated, RMI analysis shows

Nick Hedley | Climate & Energy | June 11, 2024

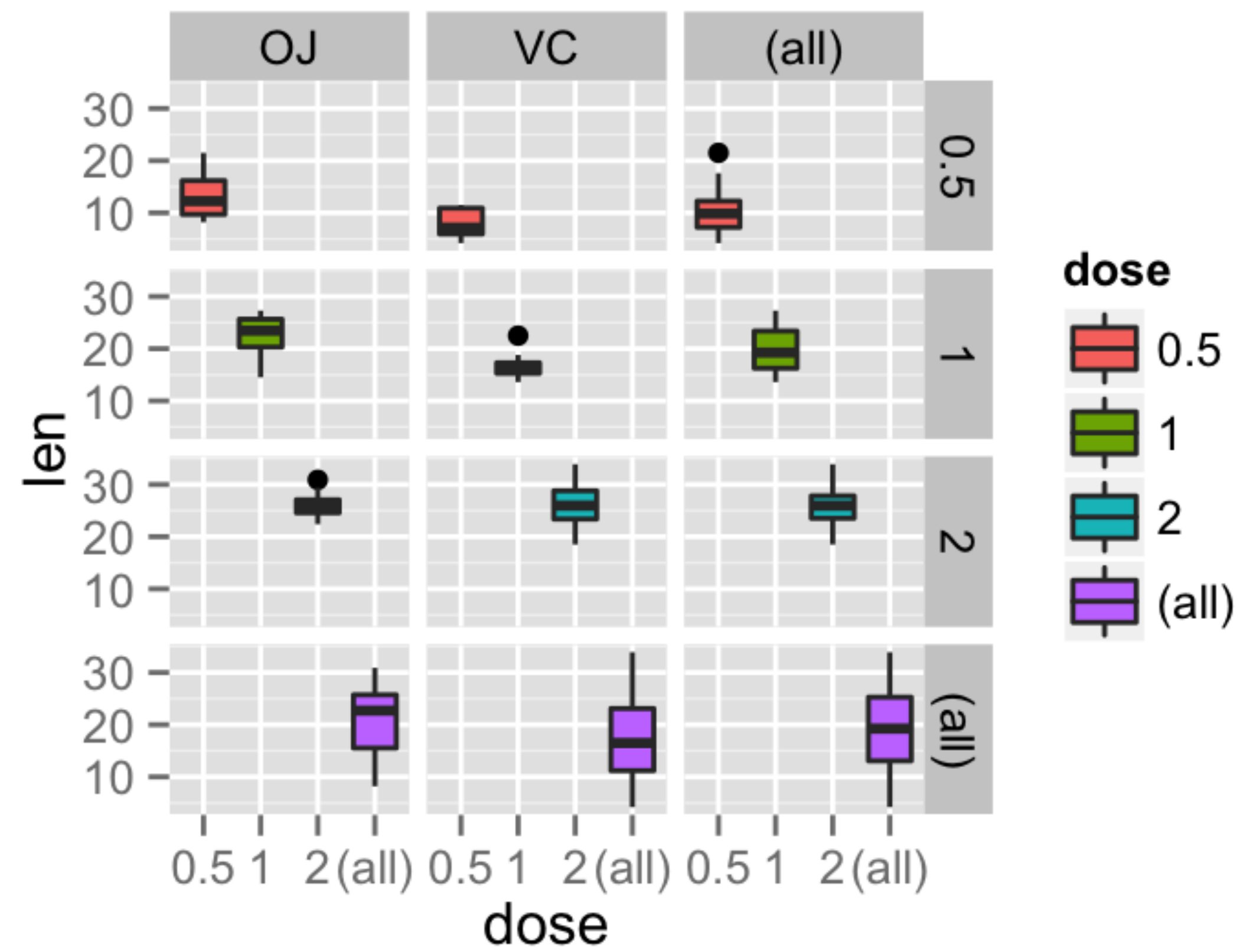


Packages

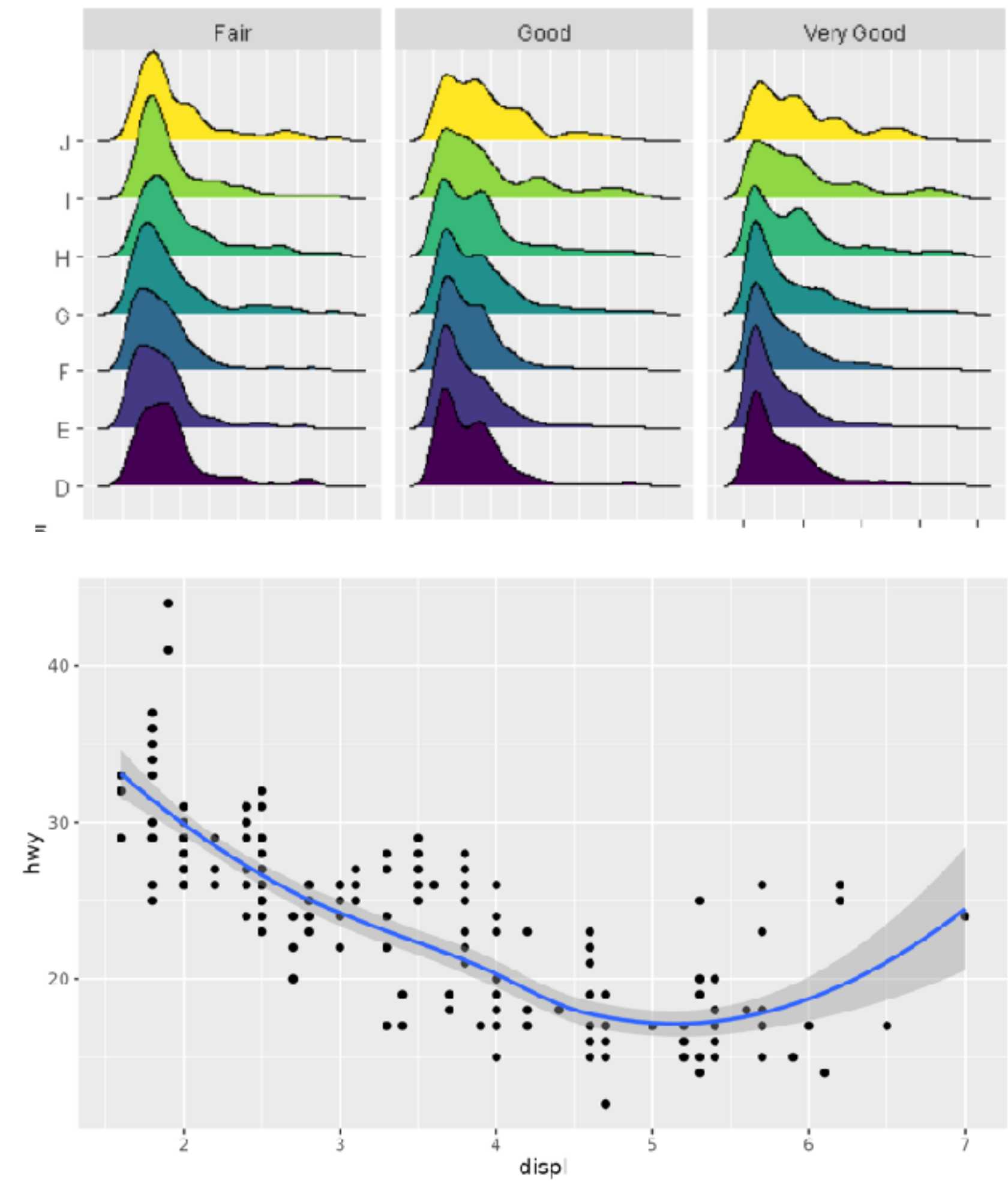
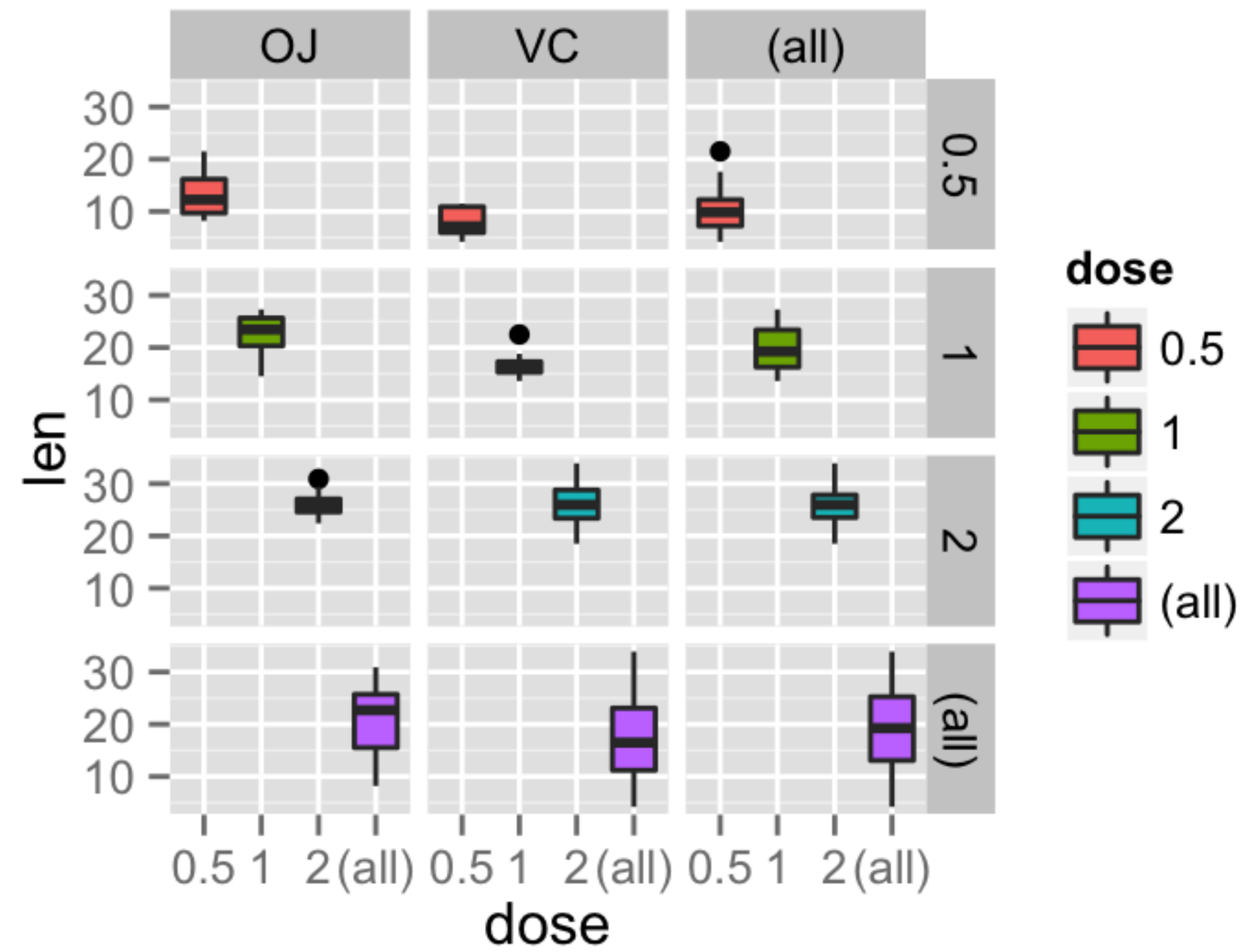


GGplot

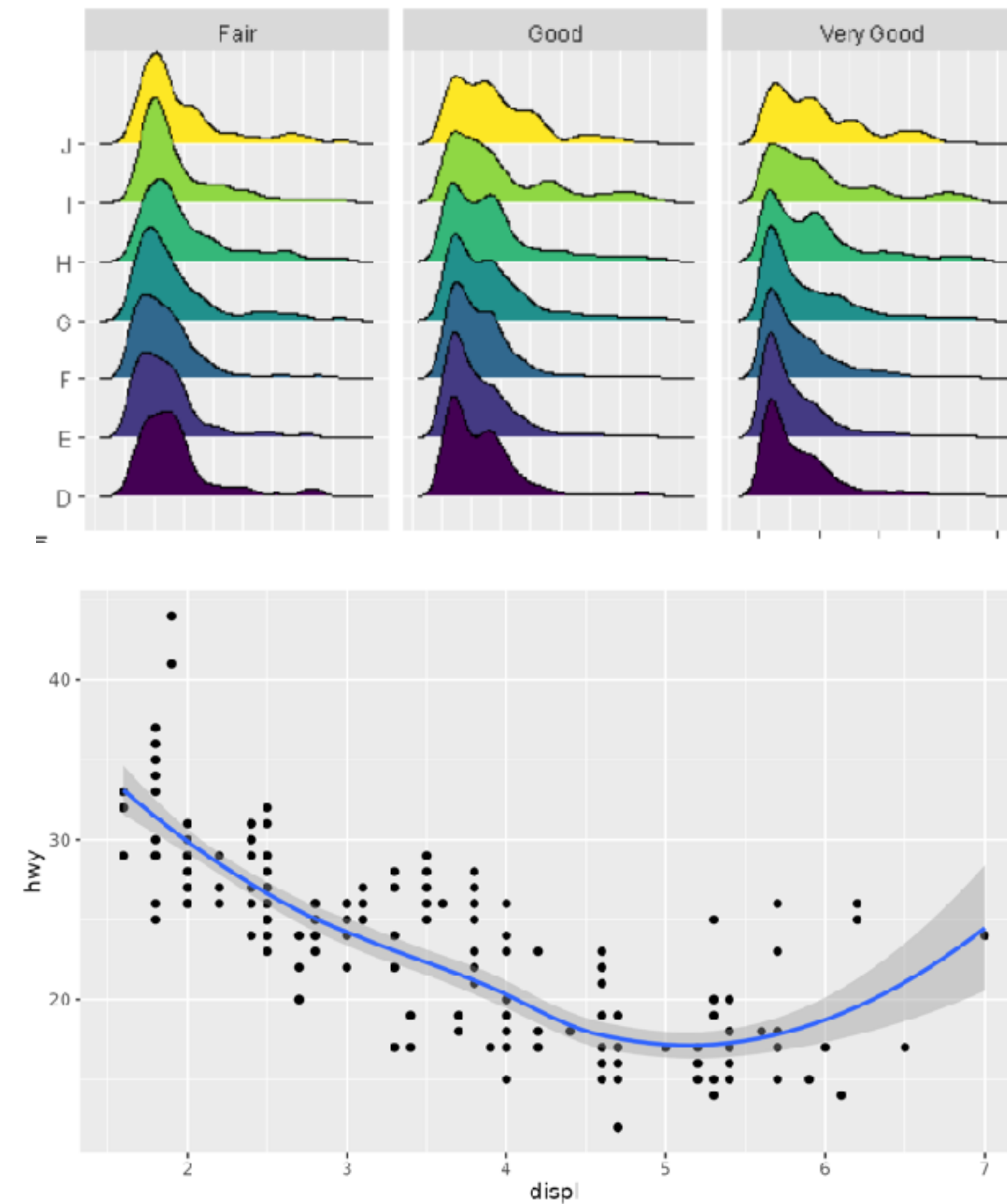
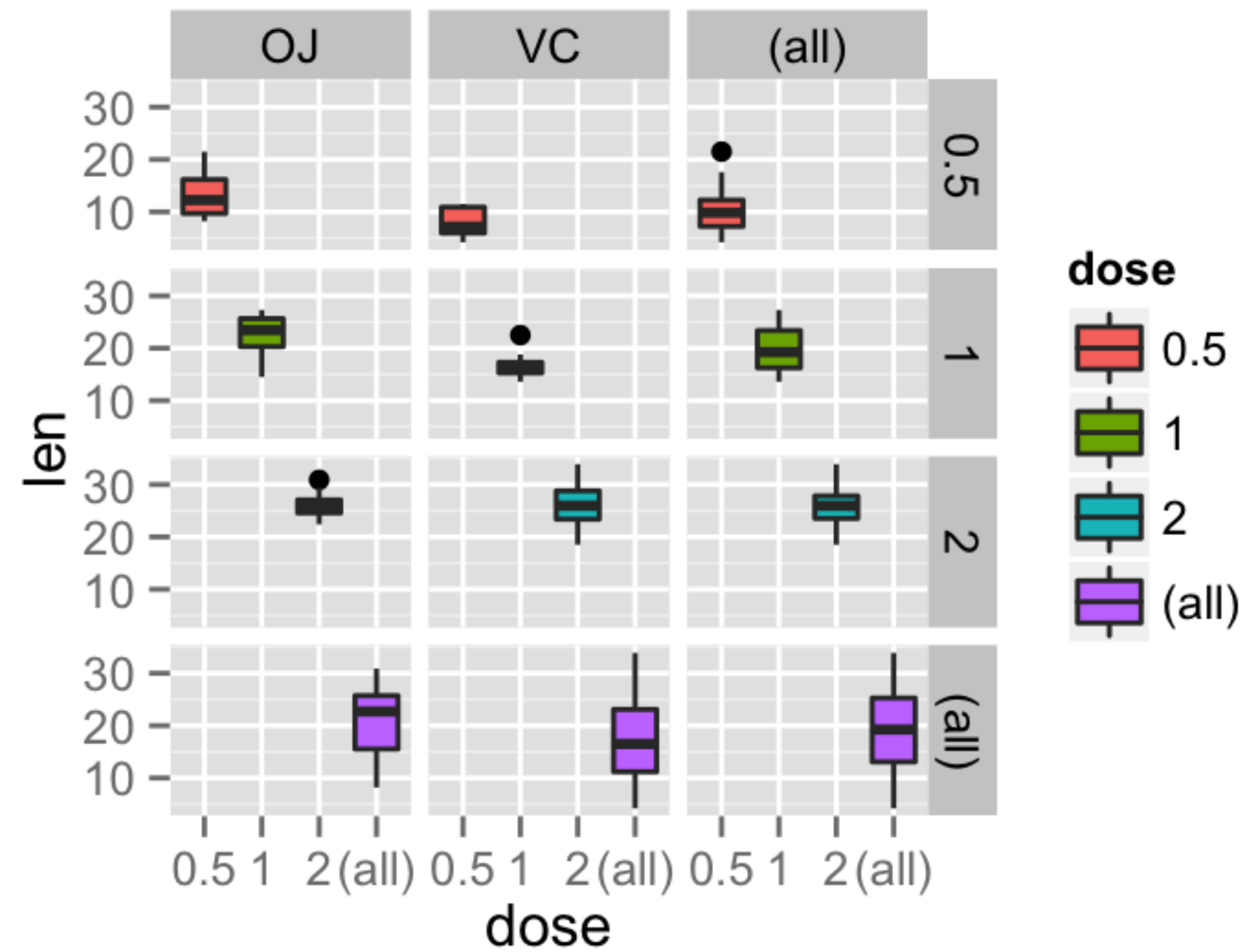
GGplot



GGplot



GGplot



GGplot kan hele mooie figuren maken, maar dat behandelen we niet in deze cursus

Soms hebben packages ook goede ‘voorbeelddata’

Soms hebben packages ook goede ‘voorbeelddata’

Oefening 2.2 (Pinguïns)



Sommige packages zijn er vooral om voorbeeld-datasets te verkrijgen. Een daarvan is

`palmerpenguins`:

```
install.packages("palmerpenguins")  
library(palmerpenguins)  
head(penguins)
```

Voer de code hierboven uit. Je hebt nu de dataframe `penguins` tot je beschikking.

a. Hoeveel soorten pinguïns staan er in de data set?

De kolom `penguins$flipper_length_mm` geeft de flipper lengtes van de pinguïns. Echter is er bij enkele metingen iets misgegaan, waardoor deze de waarde `NA` (Not Available) heeft.

b. De functies `min`, `max` en `mean` geven respectievelijk de kleinste, grootste, en gemiddelde waarde van een lijst. Bereken deze waarden voor de flipper lengtes (hint: je kunt deze functies een tweede argument `na.rm=TRUE` meegeven om de NA's te verwijderen).

c. Gebruik de functie `summary` om te berekenen voor welk percentage van de pinguïns de sekse *niet* is bepaald.