Tips to

Start with data analysis

Steps

01

Data cleaning

02

Summary statistics

03

Plotting

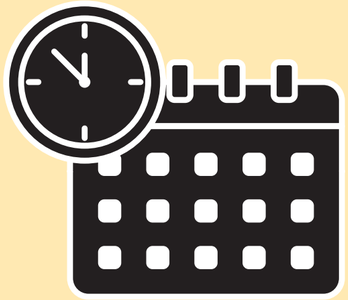
04

Drawing conclusion

Step 01

Data cleaning

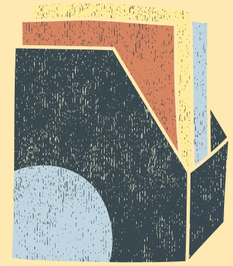
From raw data frame to data that you are confidently working with and sharing with people



Give it ~5 days

raw data

Desired data



ID	Timepoint	Germination_rate	Treatment	Other_data
9036	1	0.5	secondary	15
9036	2	0.7	primary	20
...	...	...		...

Tips

2.Cleaning it in R

- Arrange columns and variables
- Merging different data records into one dataframe
- Small checks: no NA values, no duplicated samples
- Make sure: number is number, character is character

1. Prepare not-so-bad raw data

- If raw data is good, there won't be too much work later in R
- Highly recommended to save things as csv in the end and work with csv in R

3.Sanity-check

Randomly check some data points with the original (raw) data if everything looks correct after all the merging

Step 02

Summary statistics

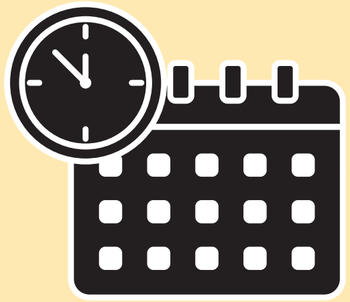
Some recommendations to start with

Overall trend of dormancy

Group data by timepoint and dormancy type, see if the overall trend is “dormancy decreases” as time increases

Looking at each genotype itself

see how dormancy increases along with different timepoints

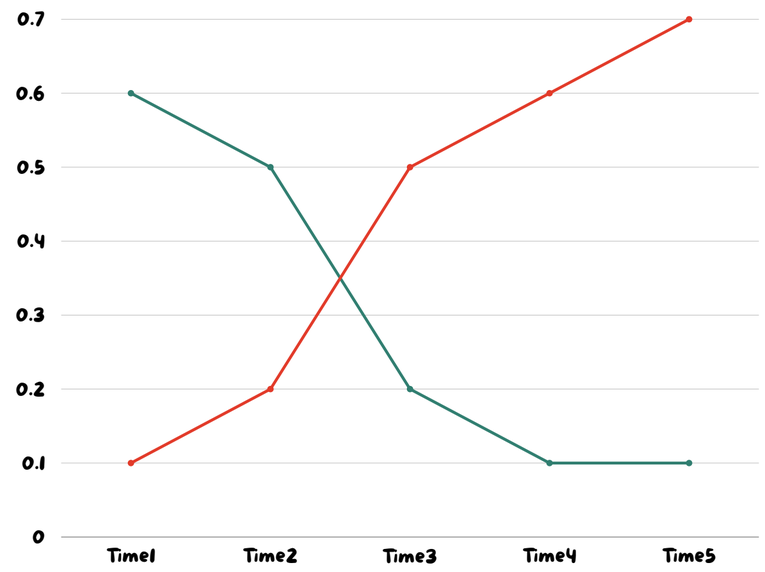


Give it ~3 days

Step 02

Summary statistics

Example plot



Some recommendations to start with

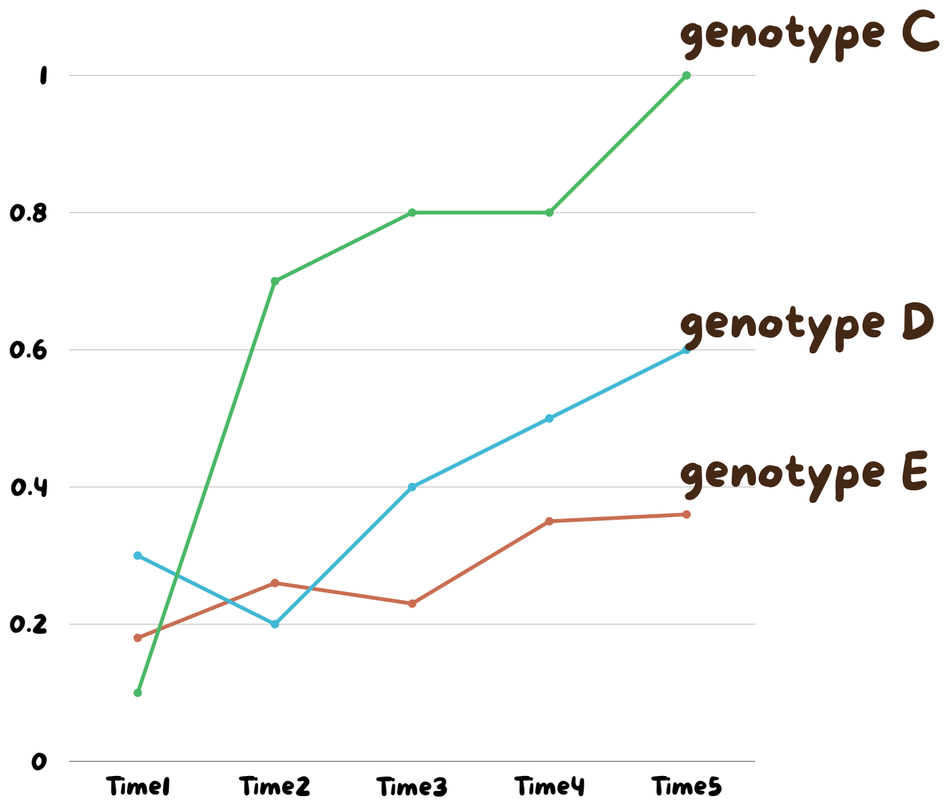
Overall trend of dormancy

Timepoint	Treatment	germination_rate
1	primary	0.2
2	secondary	0.2
1	primary	0.2
2	primary	0.2

Step 02

Summary statistics

Example plot (subset secondary dormancy)



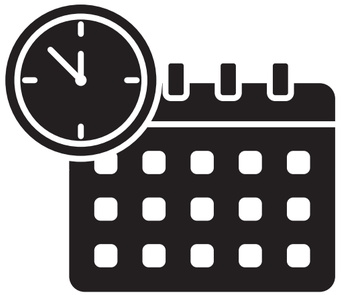
Some recommendations to start with

Looking at each genotype itself

Timepoint	Genotype	Germination_rate
1	A	0.2
2	A	0.3
3	A	0.3
1	B	0.1
2	B	0.2
3	B	0.5

03

Plotting



Give it ~5 days

Ideas

How to visualise your message of the summary statistics
Start with scatter plots, line plots, as the most common to see trends

Plot and re-plot

Present it to see if people understand what you try to show with the plot. Improve the plot or change the plot type accordingly

Tips

2.Think of how to plot in R

Maybe you need to subset the data,
arrange it in new ways, melt it, etc to plot

1.Drafting

Draft what you want to
plot on paper

3.Plotting

Finally draw your plot:)

04

Final notes

Meet up

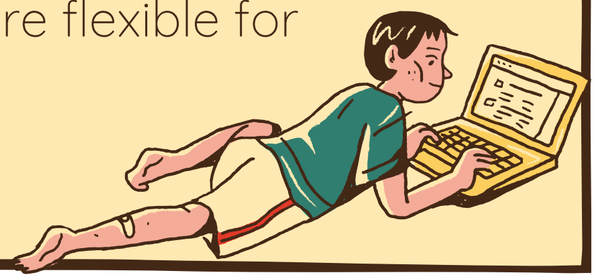
Contact me after you finish your summary. We will sit together and do statistical test where needed and discuss your data. Then meet again to review your plots:)

Need help in-process?

I'm here to help. I've been through what you're going through, ask me anything without any stress :)

My availability

I'm not available at all on 19th Dec, 22nd -25th Dec, 5th of Jan. The rest are flexible for Zoom meetings. Offcially back in office on 8th Jan.



Example of focus goals

Must-do

- Comfortable with data wrangling in R
- Summary statistics
- Know how to use ggplot

Could-do

- Work with the bigger data set
- Using R markdown
- Delve into statistics more
- Make very fancy plots

Grab a pencil or computer and...

Happy R!