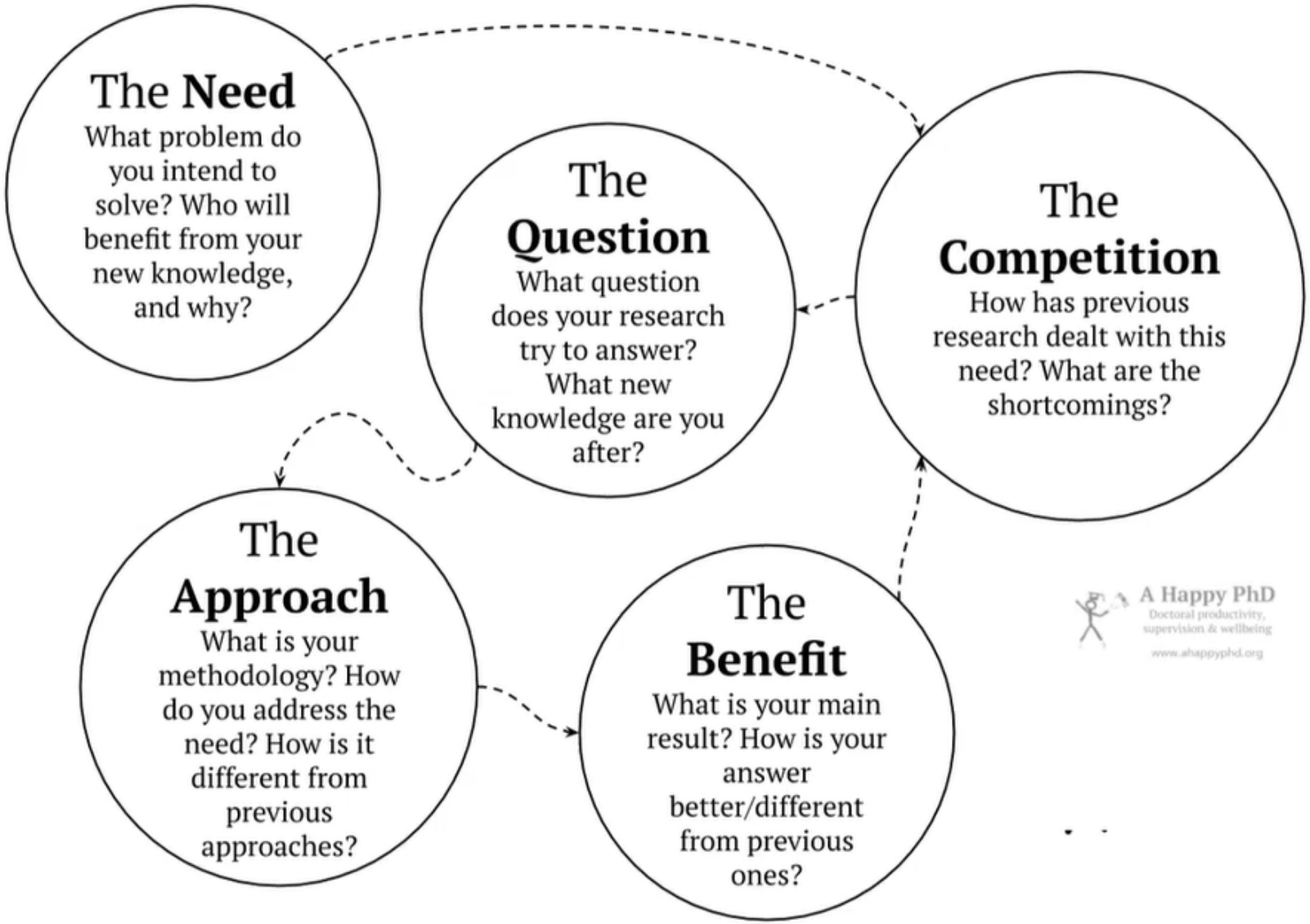
Scientific writing

Lutz Weihermüller & Linda Klamann

Often you will be blocked in writing

We all are to some extend but without start no end

Conceptualize your study: Only a good concept can be written up



In the writing process: Ask the right questions

- Why is your (sub)field important?
- What is the problem and/or knowledge gap you are addressing?
- Why does it matter?
- How did you achieve what other could not, what is novel?
- Who will care?

The scientific method to write a paper

- Ask a question
- Do Background Research
- Construct a hypothesis
- Test the hypothesis
- via experiments, models ….
- Analyse your data
- Draw Conclusions

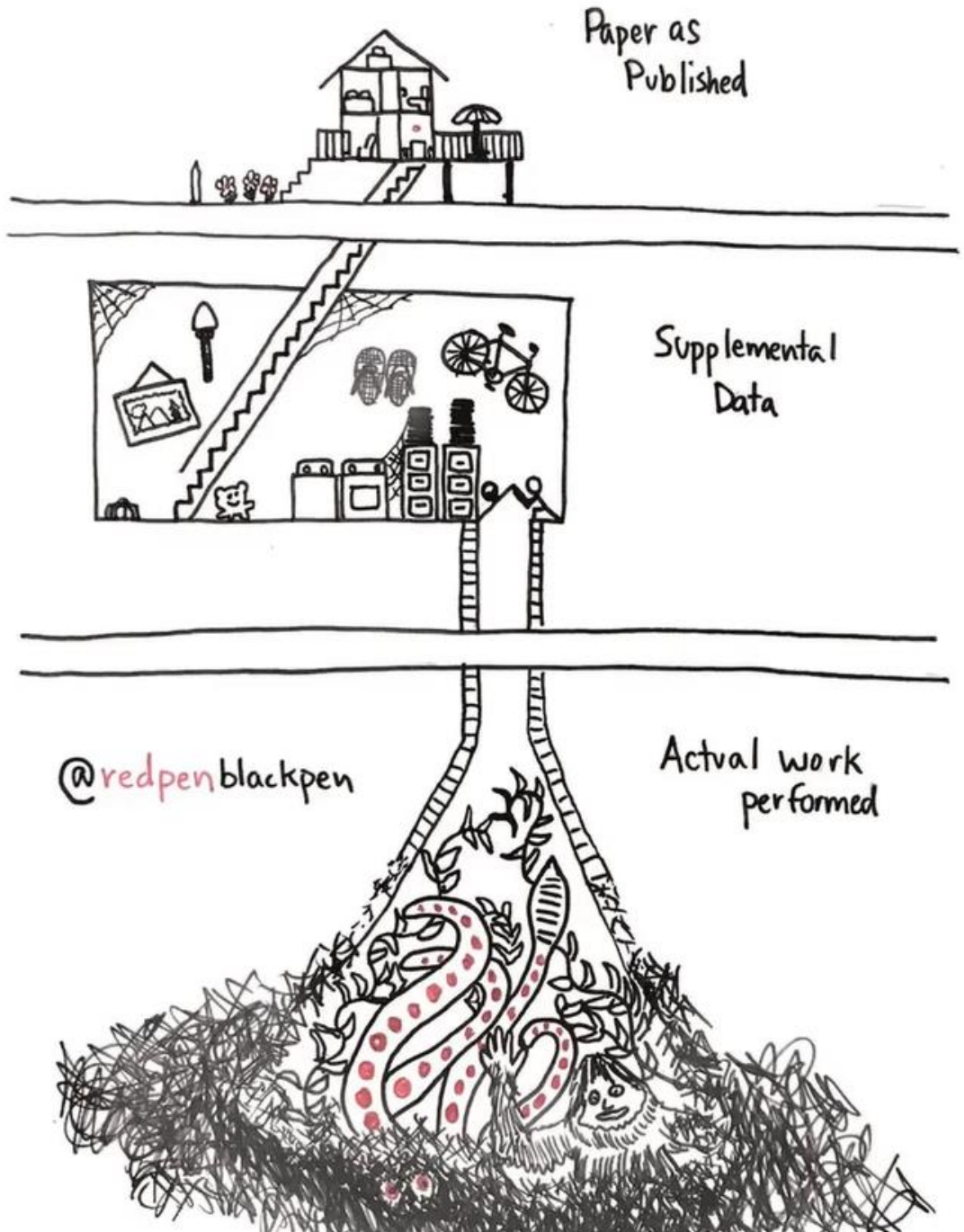
- Communicate your findings

Shaping your story

Prioritize…..

streamline your “puzzle”

…. and keep it simple



Chose a good title

Title should be short

but should be descriptive, specific, and concise, accurately reflecting the paper's content while including essential keywords for discoverability

Example for the paper title

Research question	One-sentence answer
"What is my paper about?"	*"My paper studies how program volume affects outcomes for liver transplant patients on waiting lists."*
"What methods/techniques did I use to perform my study?	*"It's a case study."*
"What or who was the subject of my study?"	*"I studied 60 cases of liver transplant patients on a waiting list throughout the US aged 20-50 years."*
"What did I find?"	*"My study revealed a positive correlation between waitlist volume and negative prognosis of transplant procedure."*

Example for the paper title

Step 2: Identify research study keywords

One-sentence answer	Title keywords/phrases
"My paper studies how program volume affects outcomes for liver transplant patients on waiting lists."	-program volume -liver transplant patients -waiting lists -outcomes
"This is a case study."	-case study
"I studied 60 cases of liver transplant patients on a waiting list throughout the US aged 20-50 years."	-US/age 20-50 -60 cases
"My study revealed a positive correlation between waitlist volume and negative prognosis of transplant procedure."	-positive correlation between waitlist volume and negative outcomes

Example for the paper title

Step 3: Research title writing: use these keywords

“We employed a case study of 60 liver transplant patients around the US aged 20-50 years to assess how the waiting list volume affects the outcomes of liver transplantation in patients; results indicate a positive correlation between increased waiting list volume and negative prognosis after the transplant procedure.”

The title above is clearly much too long for a research paper title. This is why you have to trim and polish your title

Example for the paper title

Step 4: Shorten

"We employed a case study of 60 liver transplant patients around the US aged 20-50 years to assess how waiting the list volume affects the outcomes of liver transplantation in patients; results indicate a positive correlation between increased waiting list volume and negative prognosis after the transplant procedure."

Still too long

Example for the paper title

Step 4: Shorten

"We employed A case study of 60 liver transplant patients around the US aged 20-50 years to assessing the impact of waiting the list volume affects the on outcomes of liver transplantation in patients; results indicate a positive correlation between increased waiting list volume and negative prognosis after the transplant procedure."

A case study of liver transplant patients assessing the impact of waiting list volume on outcomes of transplantation prognosis

As "case study" is a method and not a search word: Rearrange the sentence

Assessing the impact of waiting list volume on outcome and prognosis in liver transplantation patients

Example for the paper title

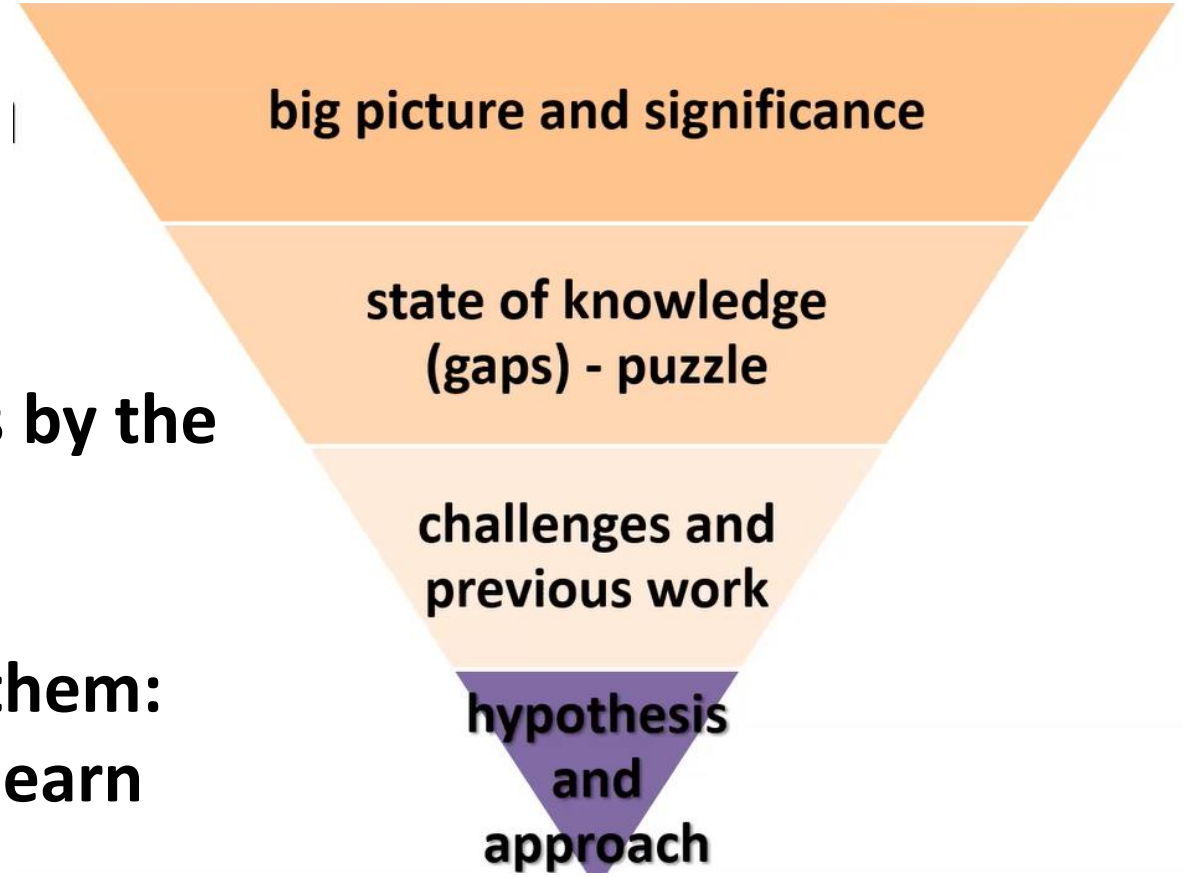
Important keyword for online search

Assessing the impact of waiting list volume on outcome and prognosis in liver transplantation patients

Selection of possible titles

The Introduction

- This is where you can take your readers by the hand
- Gently guide them to where you want them:
  Excited about research and eager to learn more what you have found out



Materials and Method

Provide all information about methods and materials used

Field description
Measurements
Theory (e.g. for models or workflows)
Statistical methods used

Materials and Method

Provide all information about methods and materials used

Field description
Measurements
Theory (e.g. for models or workflows)
Statistical methods used

Be detailed but to not bee to long

Write cook book style

and separate parts from each other (subsections)

Materials and Method

Provide all information about methods and materials used

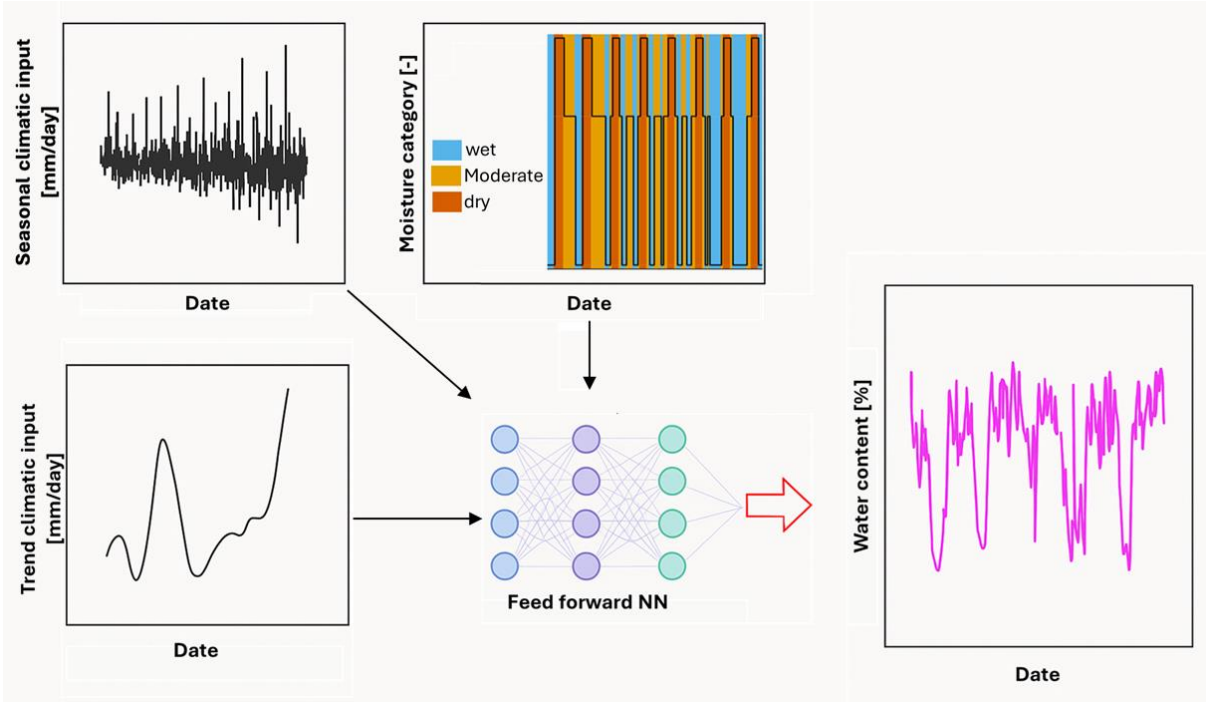
Field description
Measurements
Theory (e.g. for models or workflows)
Statistical methods used

Do not refer only to other papers in the methods such as: *The measurements have been performed as shown by Wang et al. (2012)*. Give at least brief information and maybe show details in the Supplementary

Materials and Method

Provide all information about methods and materials used

Field description
Measurements
Theory (e.g. for models or workflows)
Statistical methods used



Results and Discussion

Sort the chaos

Start to describe the figures

- what do you see
- what is unexpected
- where do values differ between treatments/soil etc.

Define key messages / key messages (rule of thumb ~3-5 messages per paper)

- What is new
- What has not been expected
- What is in line with other work

Think about how to present the messages / key messages

- They are the headlines for the discussion
  (e.g., due to limited bioaccessability of a pesticide the microbial activity has not been decreased)

Key message drives it all

- When organizing section, paragraphs and sentences,
  start with what is the most important and add necessary information after

- In other words, start with the exciting stuff, make your case, and defend your findings to
  convince the reader that your excitement is justified

Do not open battlefields for the reviewer

- If you are imprecise
- If you are inconsistent
- If you use inappropriate plots
- If you are too lengthy
- If you are too short
- If your language is weak
- If you are boring

Imprecise statements

- **Avoid statements in the introduction such as**

Stefens et al. (2012) analysed the occurrence of pesticides in Vietnam and the data of Minh (2019) showed the same problematics.

- **This is not concrete and will raise questions**

Better to be detailed

Stefens et al. (2012) analysed the occurrence of pesticides in the MKD of Vietnam and found various pesticide residues in rice field soils. The same findings have been made by Minh (2019) who studied 10 field in the region.

Inconsistency

Now Linda takes over

Inconsistency

Inconsistencies can be either in the text or between text and figures/tables

The highest measured wheat biomass of 8 t/ha was archived when the volumetric water content was kept constant at 0.3 cm^3 cm^{-3}. Therefore, the soil moisture content should be kept below saturation at field capacity.

Inconsistency

Inconsistencies can be either in the text or between text and figures/tables

The highest measured wheat biomass of 8 t/ha was archived when the volumetric water content was kept constant at 0.3 cm^3 cm^{-3}. Therefore, the soil moisture content should be kept below saturation at field capacity.

Do you find the second inconsistency ?

Inconsistency

Inconsistencies can be either in the text or between text and figures/tables

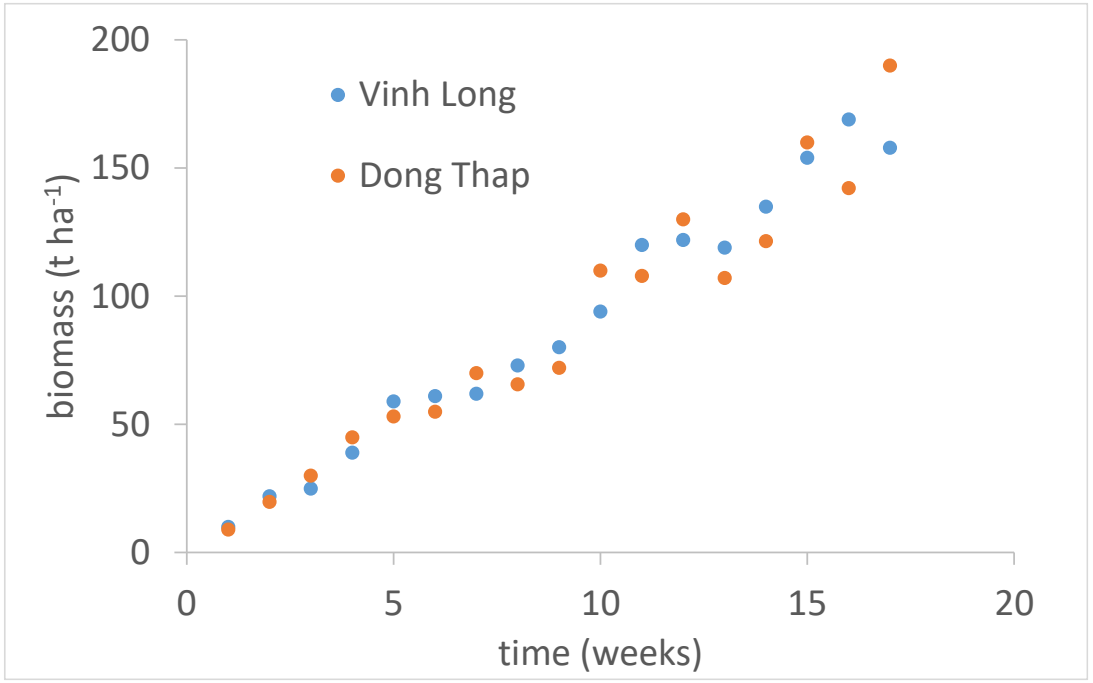
The highest measured wheat biomass of 8 t/ha was archived when the volumetric water content was kept constant at 0.3 cm^3 cm^{-3}. Therefore, the soil moisture content should be kept below saturation at field capacity.

The highest measured wheat biomass of 8 t/ha was archived when the volumetric water content was kept constant at 0.3 cm^3 cm^{-3}. Therefore, the soil moisture content should be kept below saturation at field capacity.

Inconsistency

Inconsistencies can be either in the text or between text and figures/tables

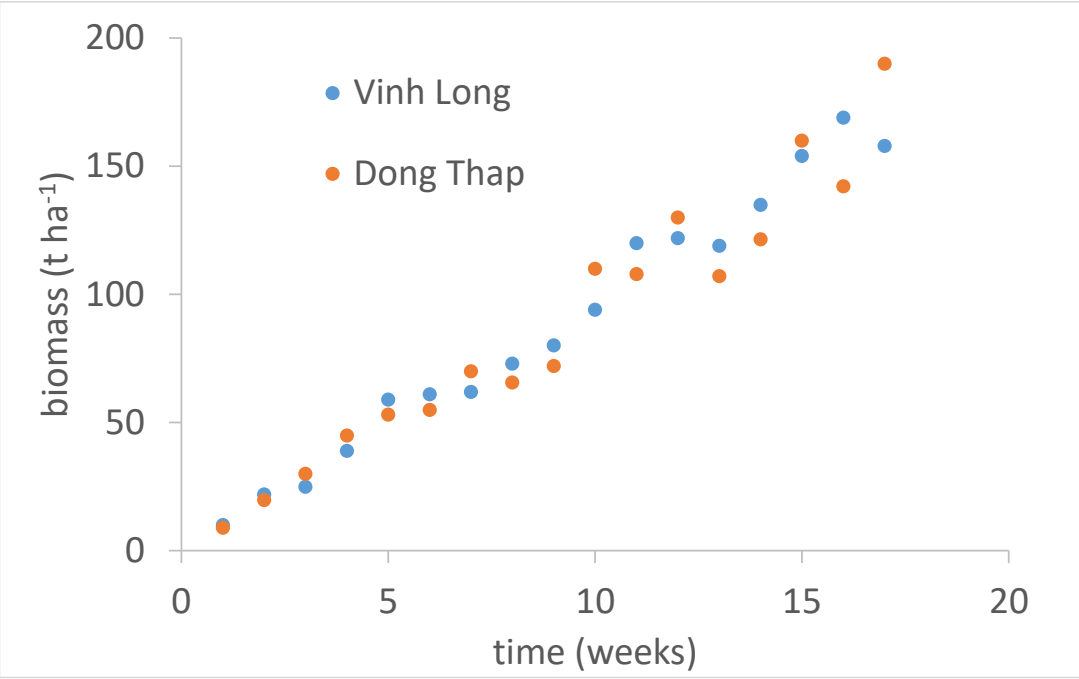
The highest measured wheat biomass of 8 t/ha was archived when the volumetric water content was kept constant at 0.3 cm^3 cm^{-3}. Therefore, the soil moisture content should be kept below saturation at field capacity.



Inconsistency

Inconsistencies can be either in the text or between text and figures/tables

The highest measured wheat biomass of 8 t/ha was archived when the volumetric water content was kept constant at 0.3 cm^3 cm^{-3}. Therefore, the soil moisture content should be kept below saturation at field capacity.

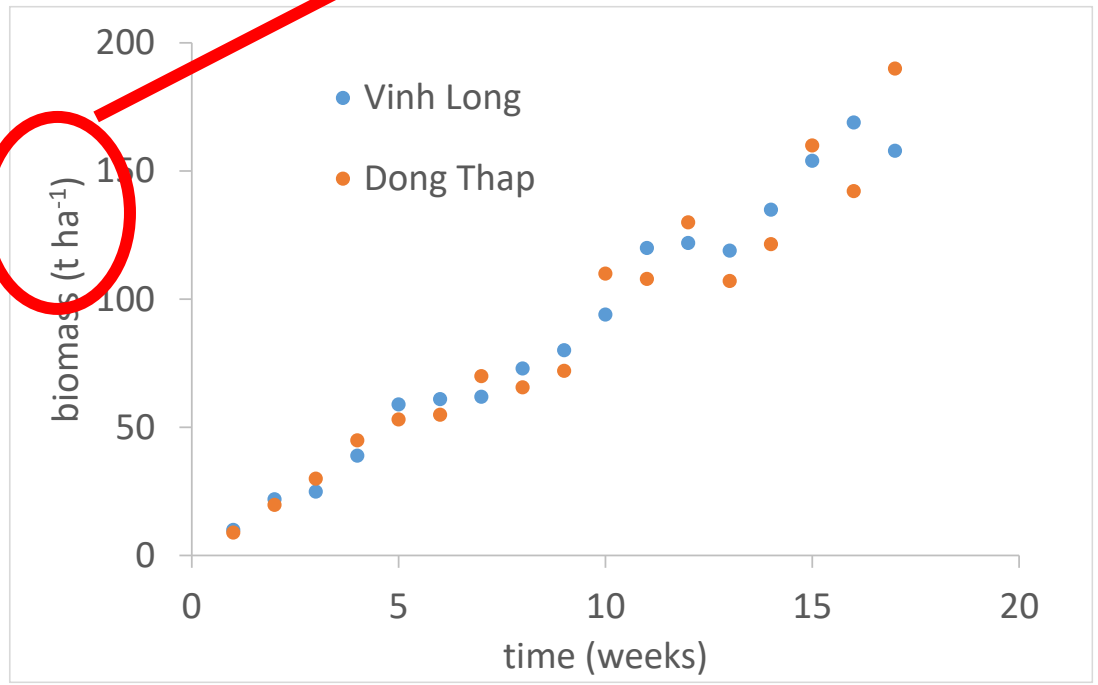


Do you find the inconsistency ?

Inconsistency

Inconsistencies can be either in the text or between text and figures/tables

The highest measured wheat biomass of 8 t/ha was archived when the volumetric water content was kept constant at 0.3 cm^3 cm^{-3}. Therefore, the soil moisture content should be kept below saturation at field capacity.



Use short sentences

Short sentences are beautiful

Multi-word phases that can be condensed

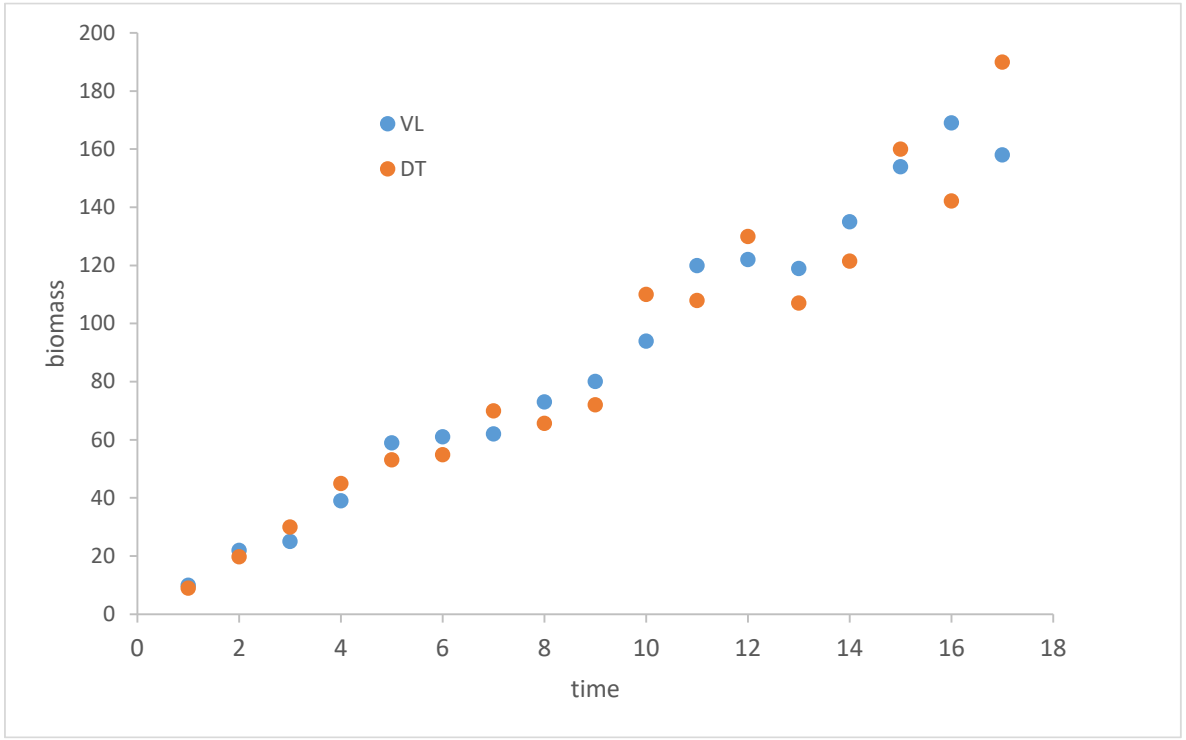
- In order to > To
- As well as > And
- Both of them > Both
- Was found to be > Was
- Substantially more > Greater
- In spite of the fact > Although
- There is no doubt that > Clearly
- Around the world > Globally or global
- A recent study has shown > A recent study showed
- In the paper, the conclusion is that > We concluded that

➢ **Or simple split long sentences into 2** (we call them spaghetti-sentences)

Inappropriate plots or tables

Often labels are too small

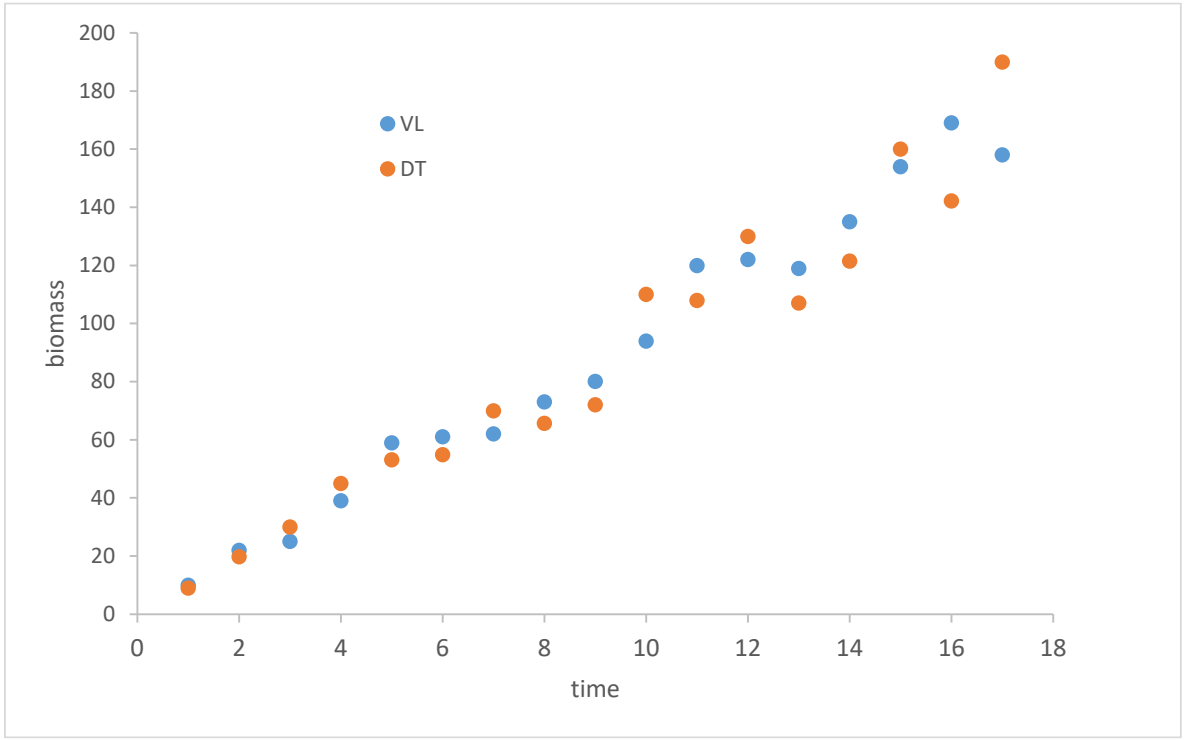
not appropriate



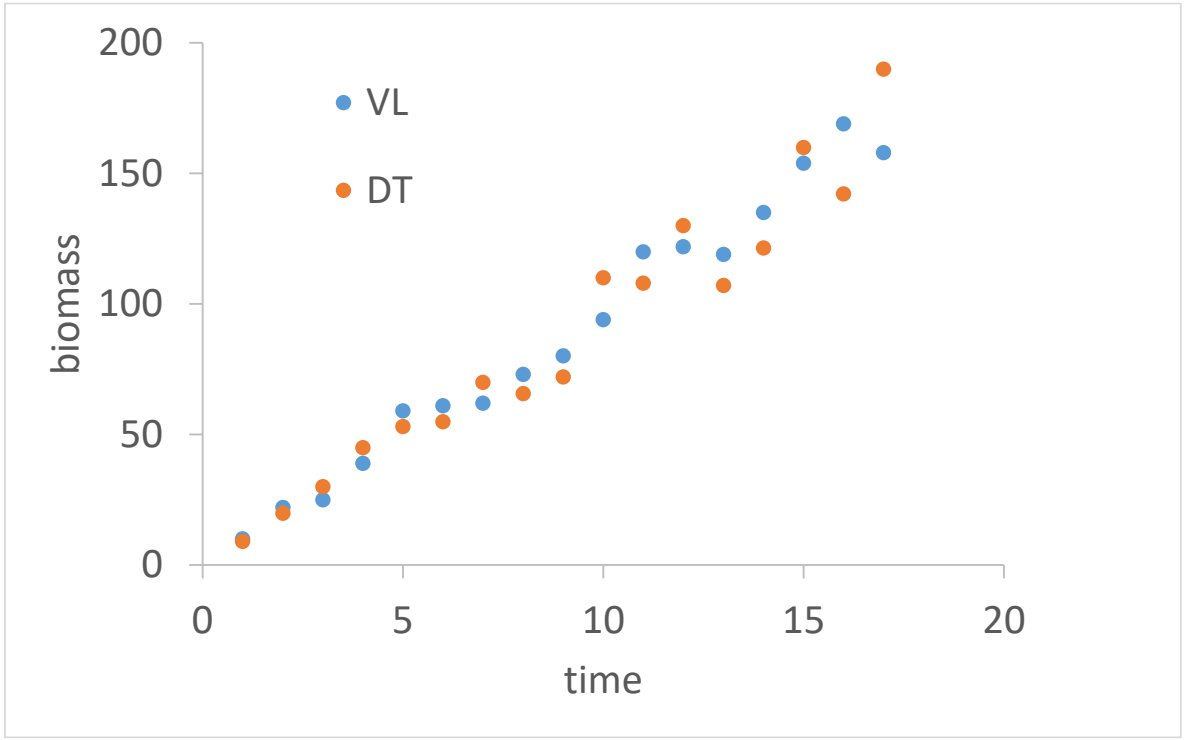
Inappropriate plots or tables

Often labels are too small

not appropriate



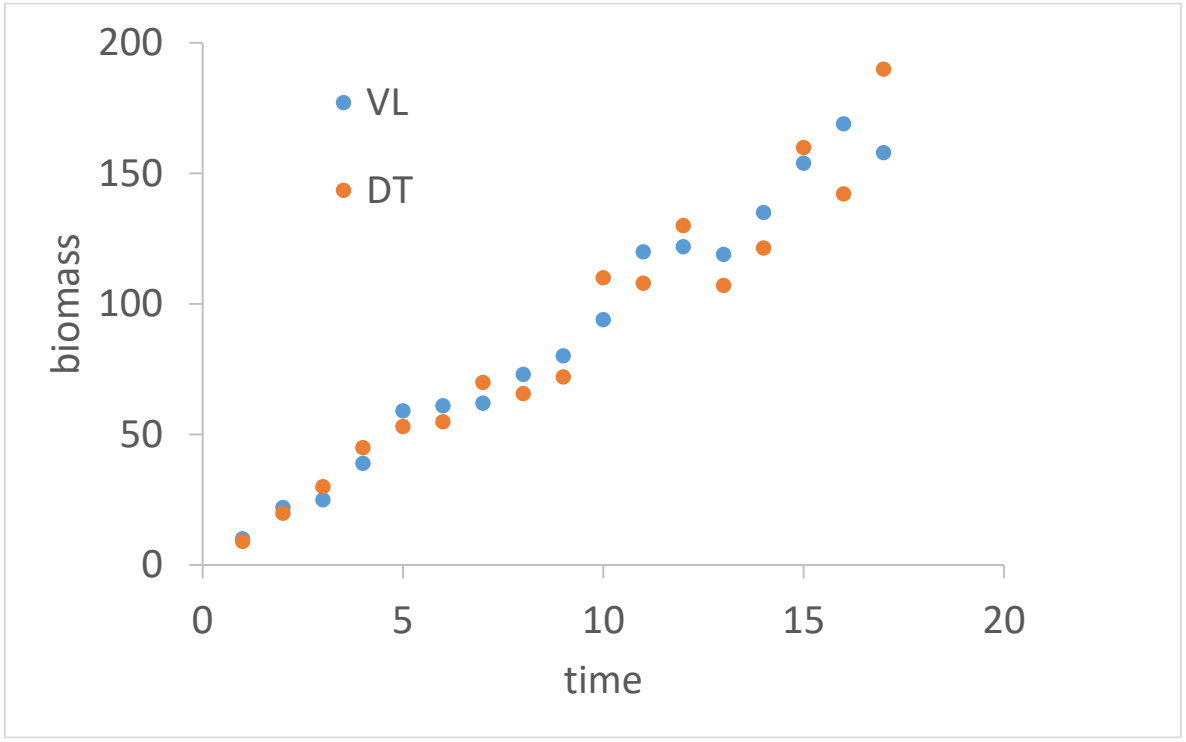
appropriate



Inappropriate plots or tables

Often units are missing

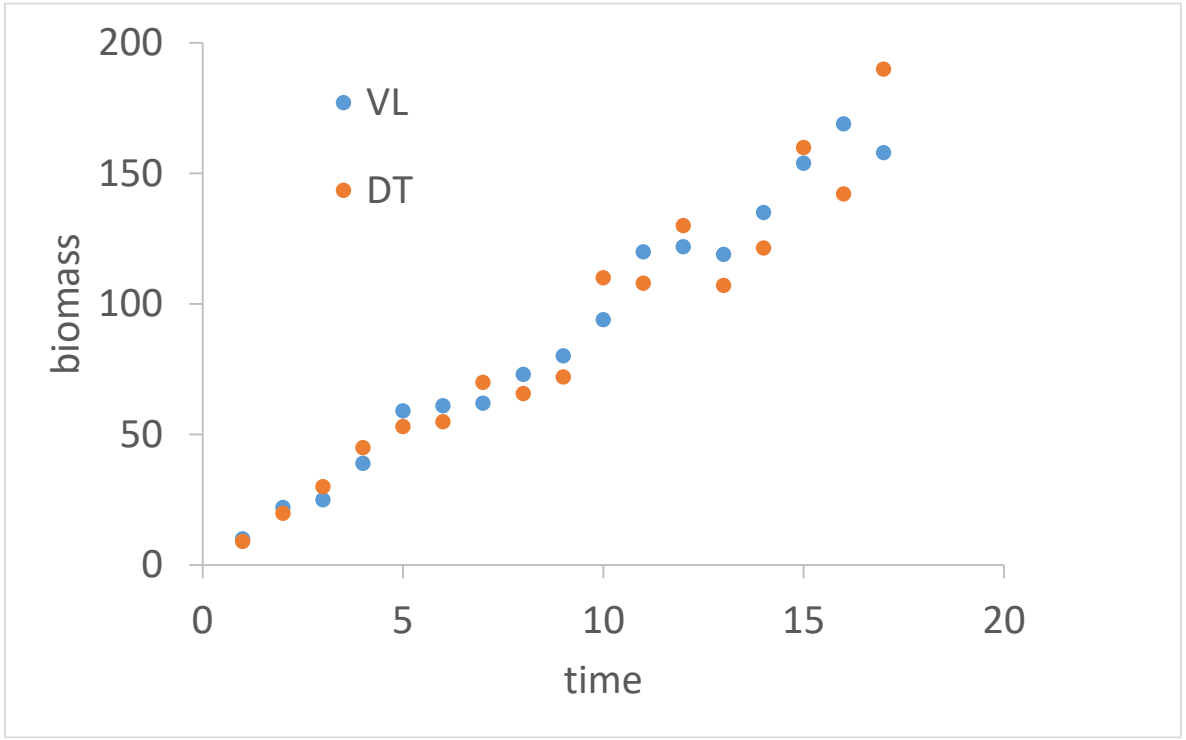
not appropriate



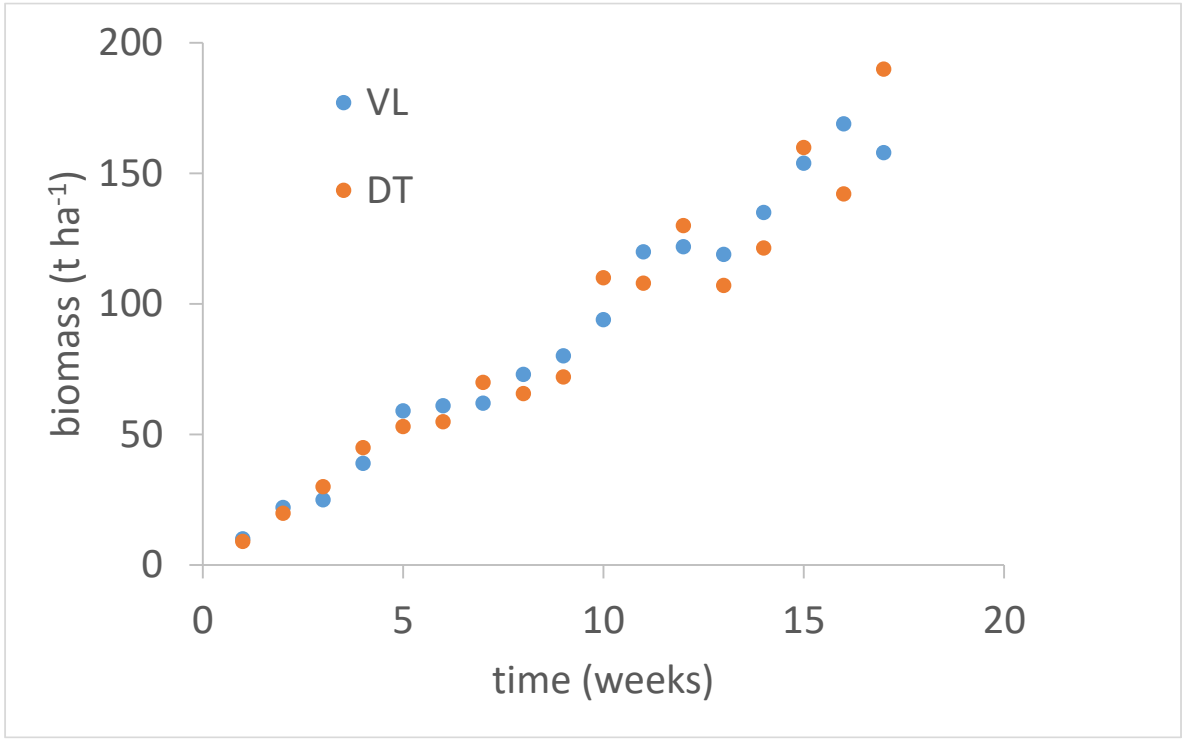
Inappropriate plots or tables

Often units are missing

not appropriate



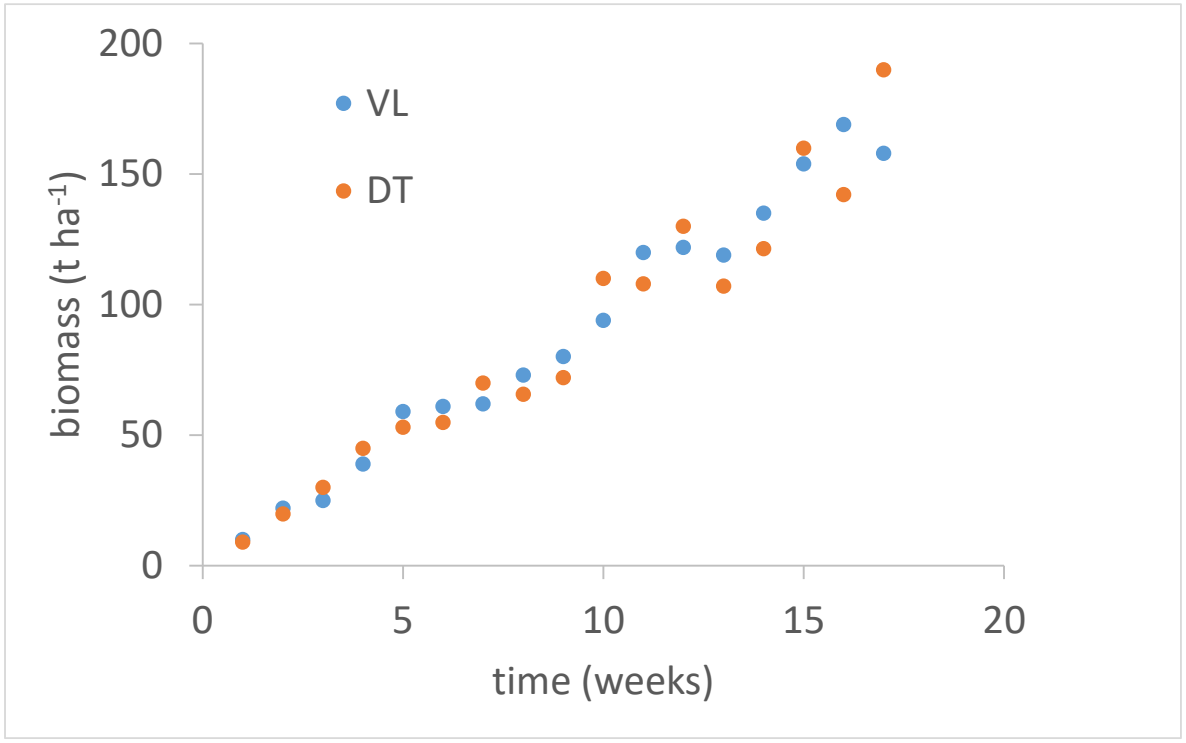
appropriate



Inappropriate plots or tables

Often legend is abbreviated

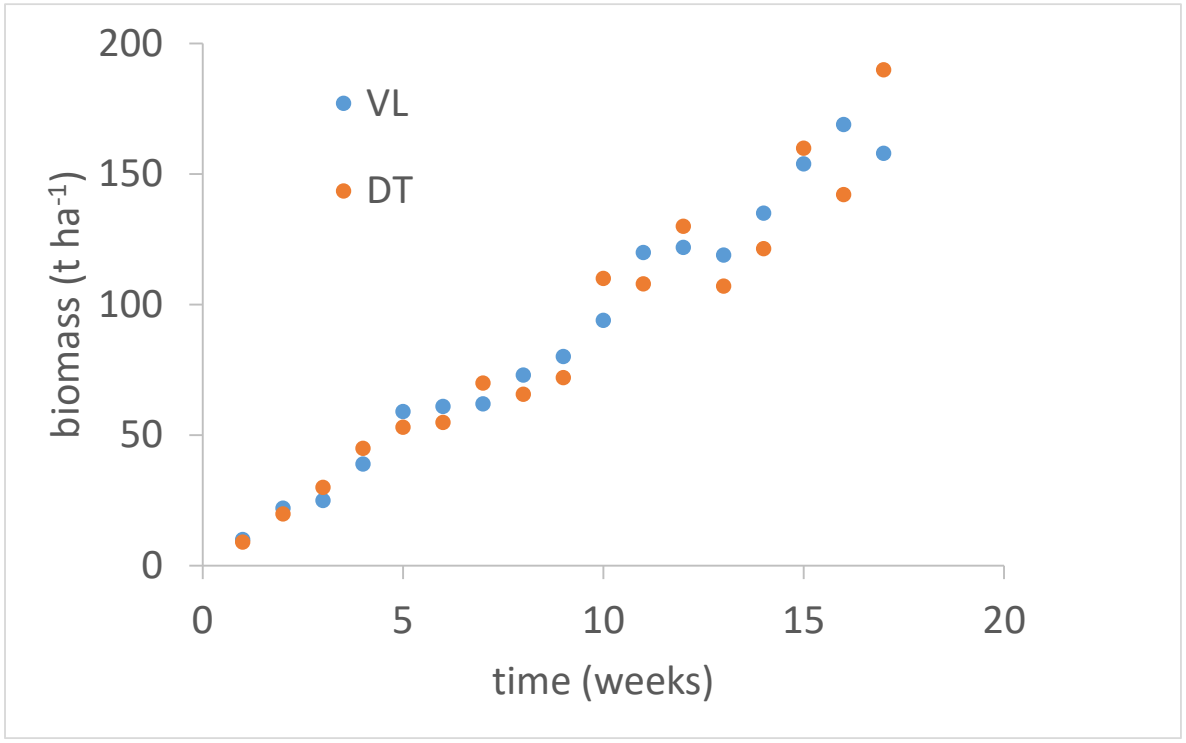
not appropriate



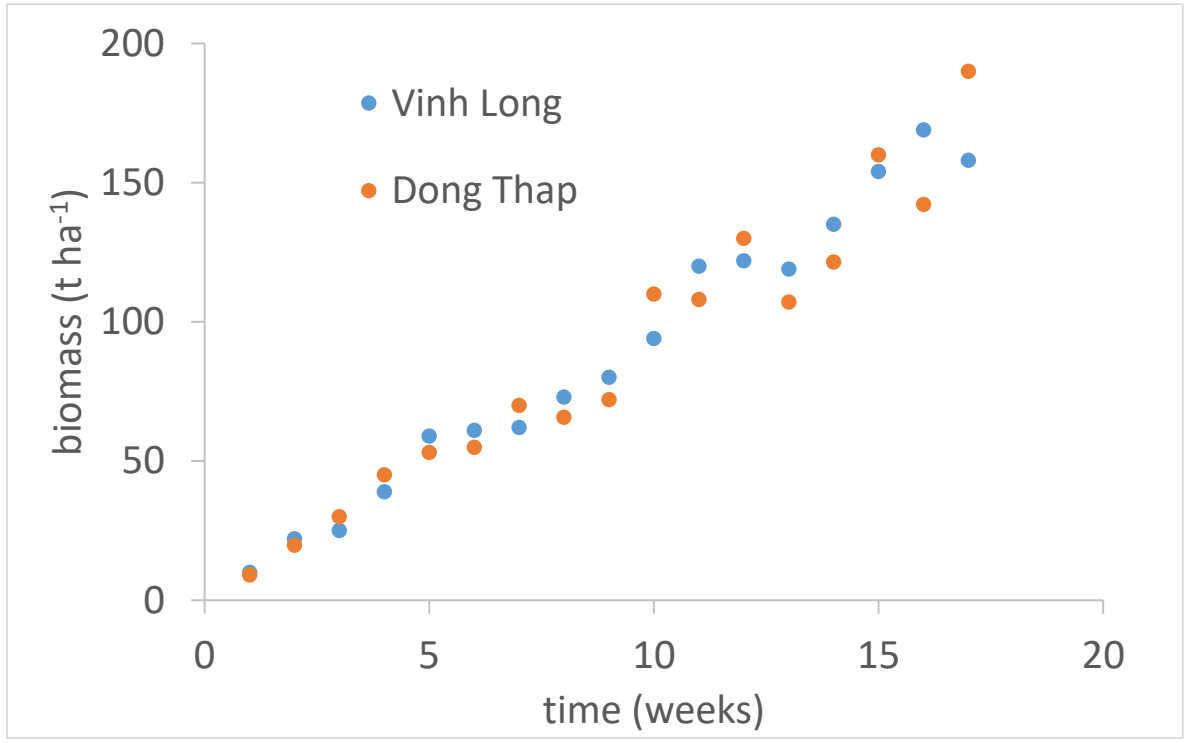
Inappropriate plots or tables

Often legend is abbreviated

not appropriate



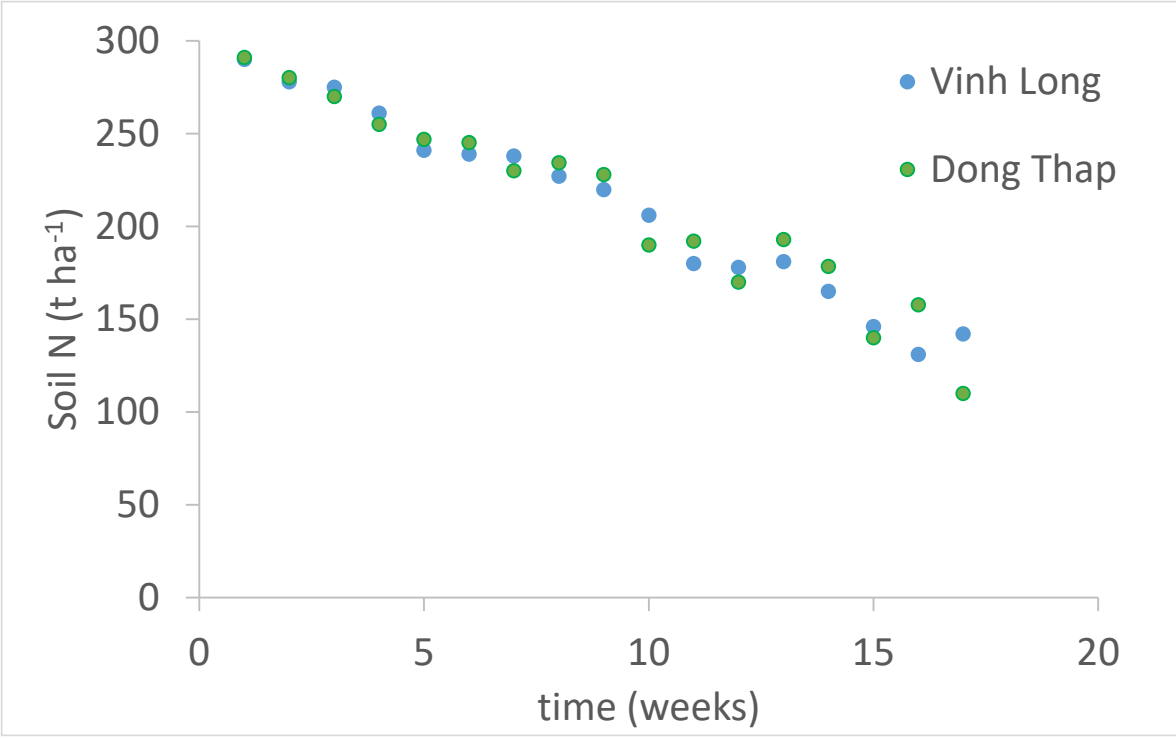
appropriate



Inappropriate plots or tables

Often colours are not easy to identify

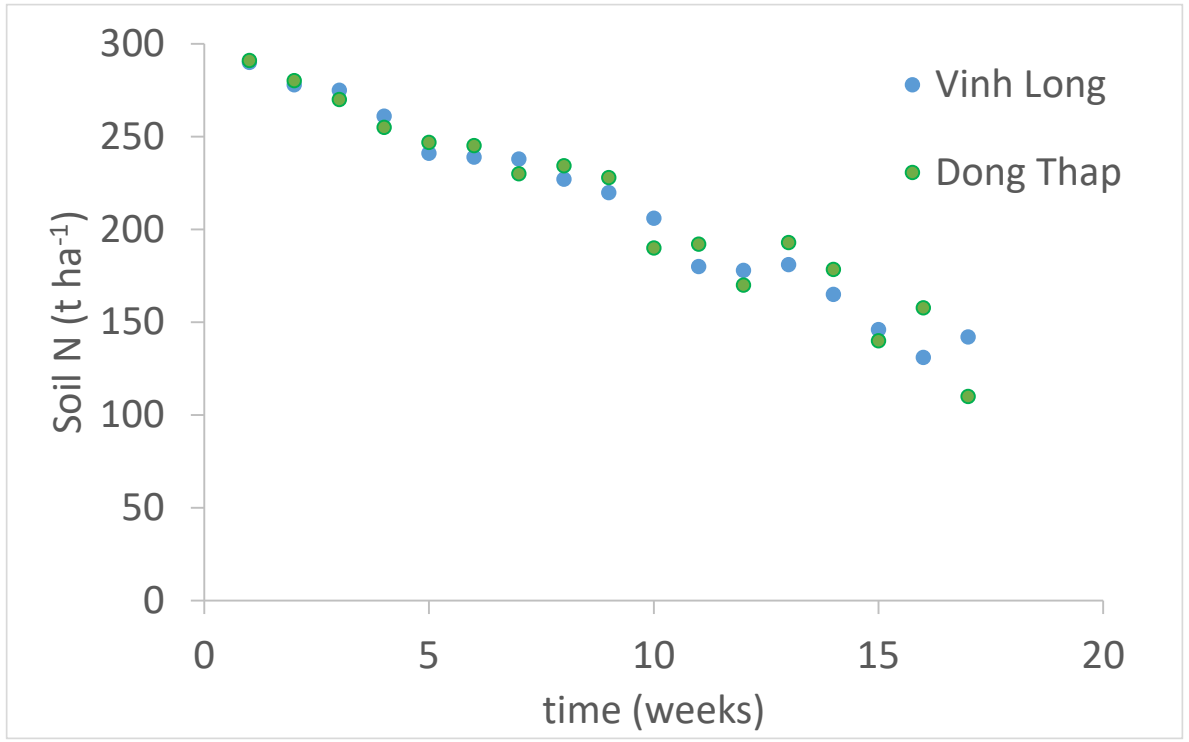
not appropriate



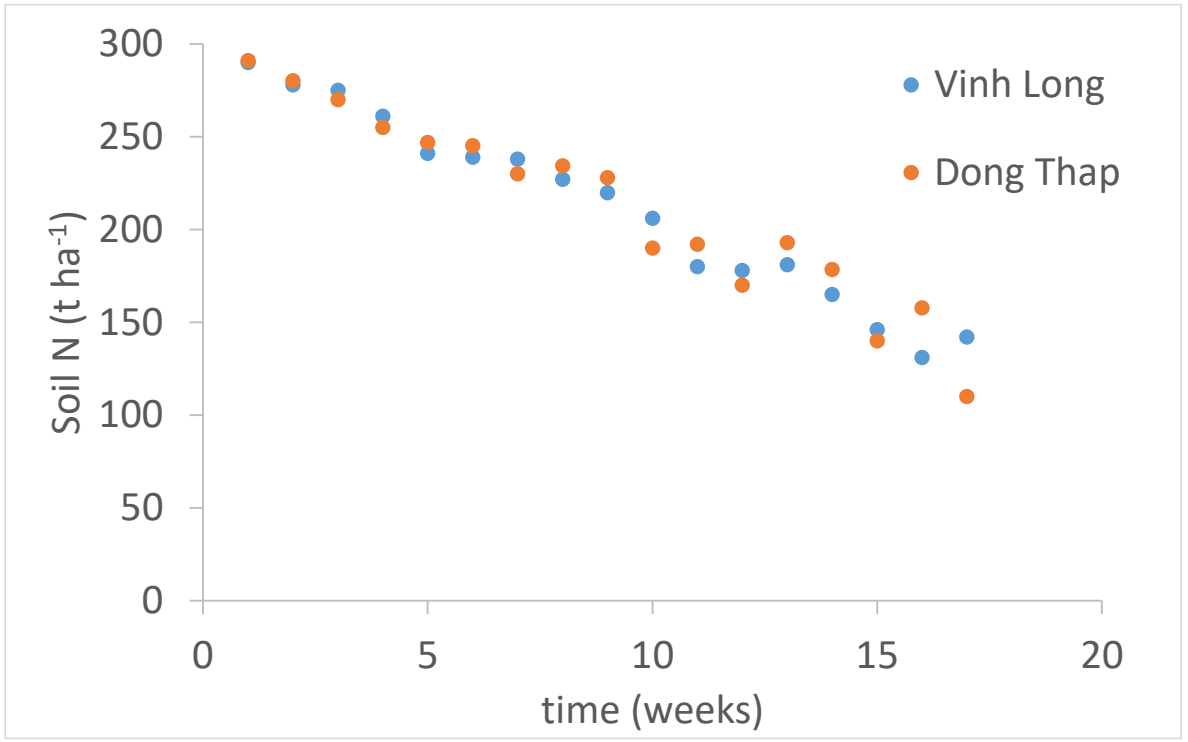
Inappropriate plots or tables

Often colours are not easy to identify

not appropriate

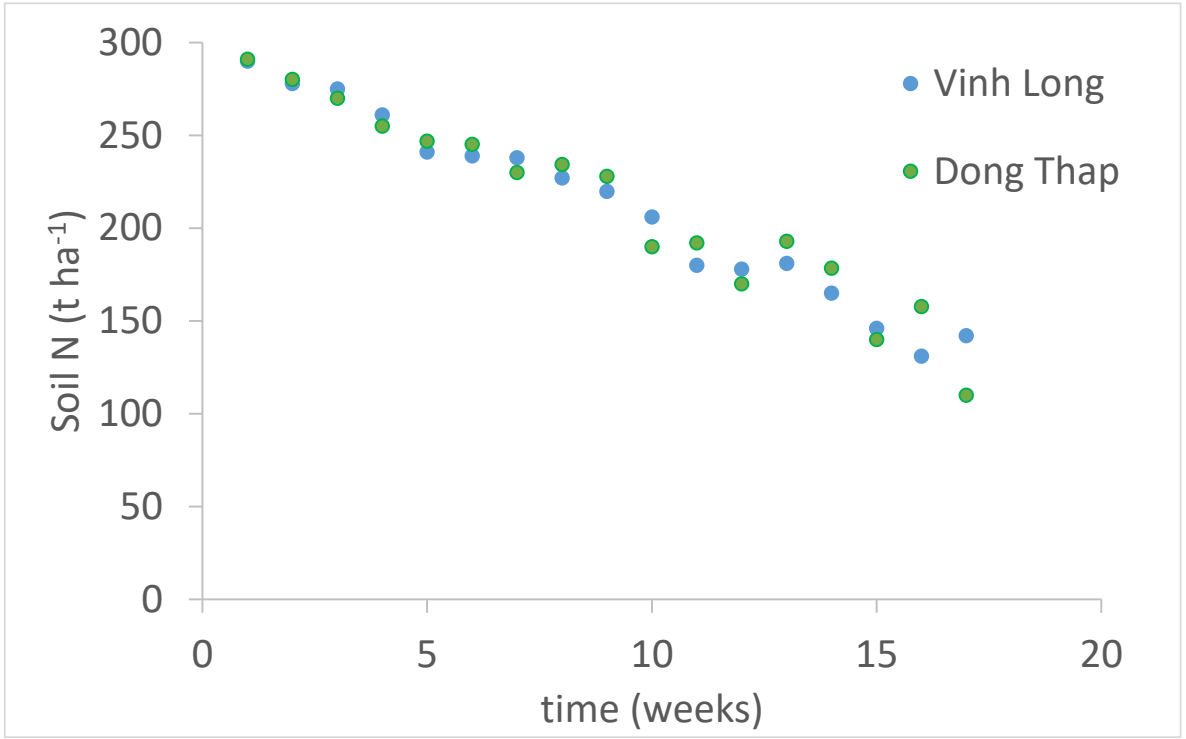


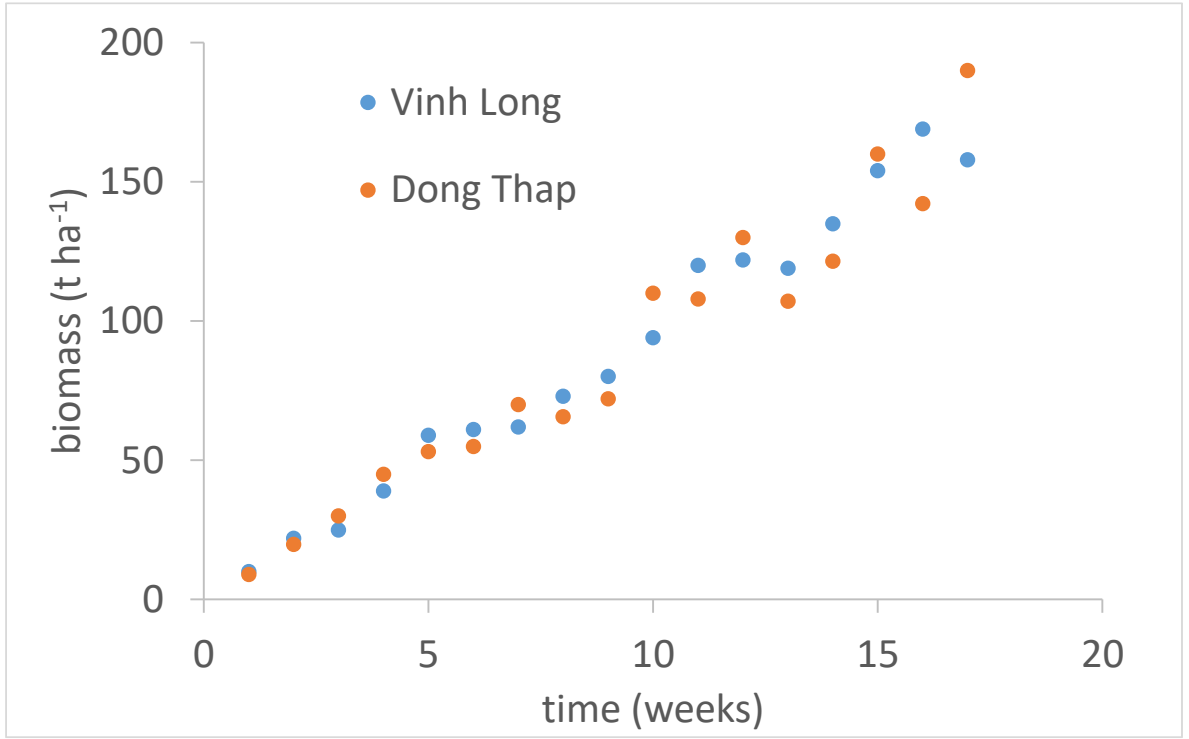
appropriate



Inappropriate plots or tables

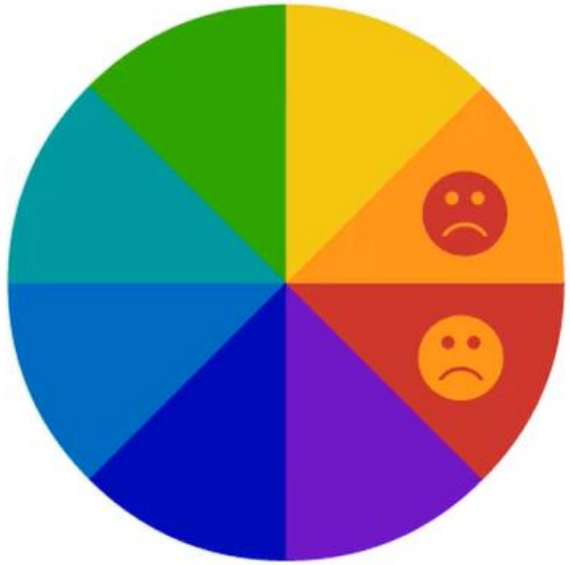
Often colours are mixed up

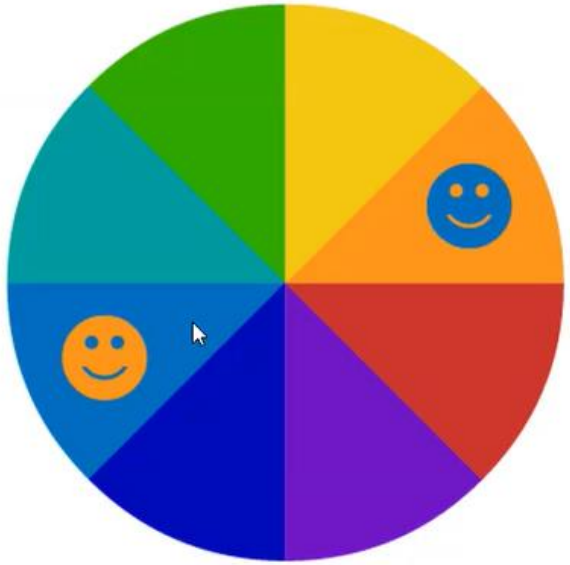




Inappropriate plots or tables

Chose colours from opposing sides of the colour wheel for best visualization (and colour blind audience)



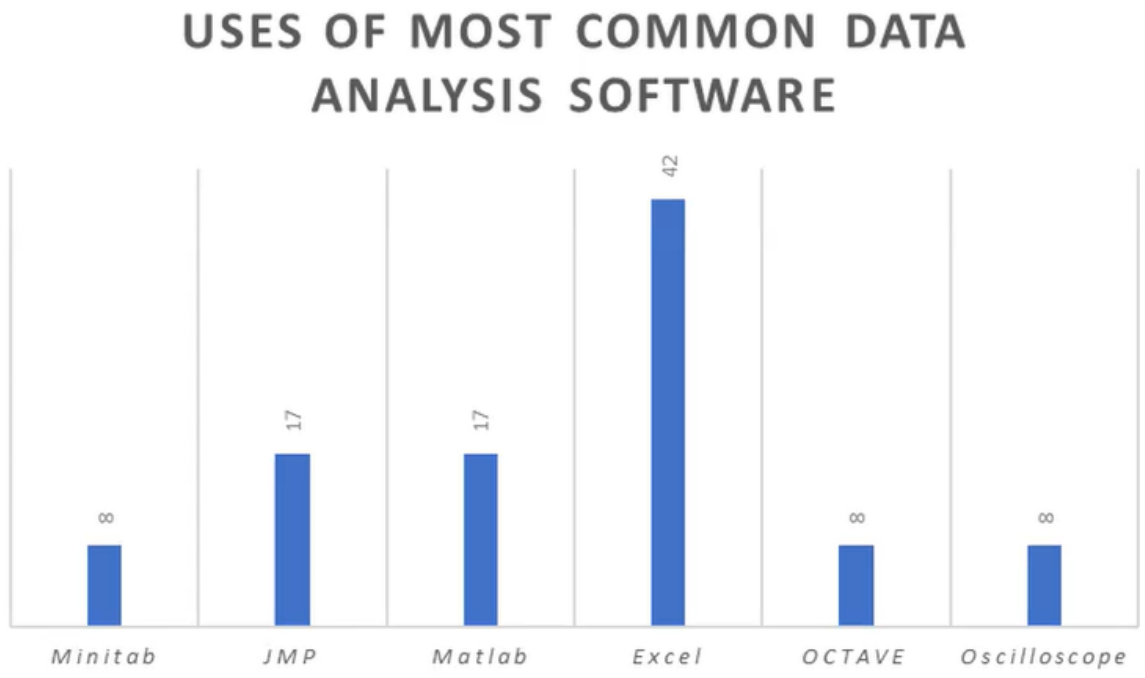


Example from https://www.simplifiedsciencepublishing.com/resources/best-color-palettes-for-scientific-figures-and-data-visualizations

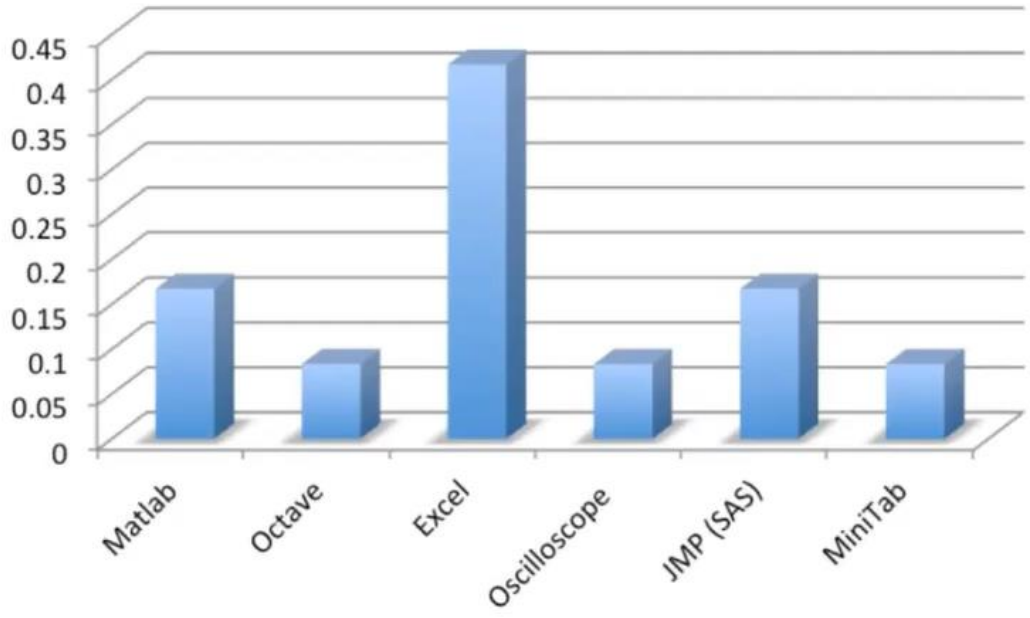
Inappropriate plots or tables

Another example

not appropriate



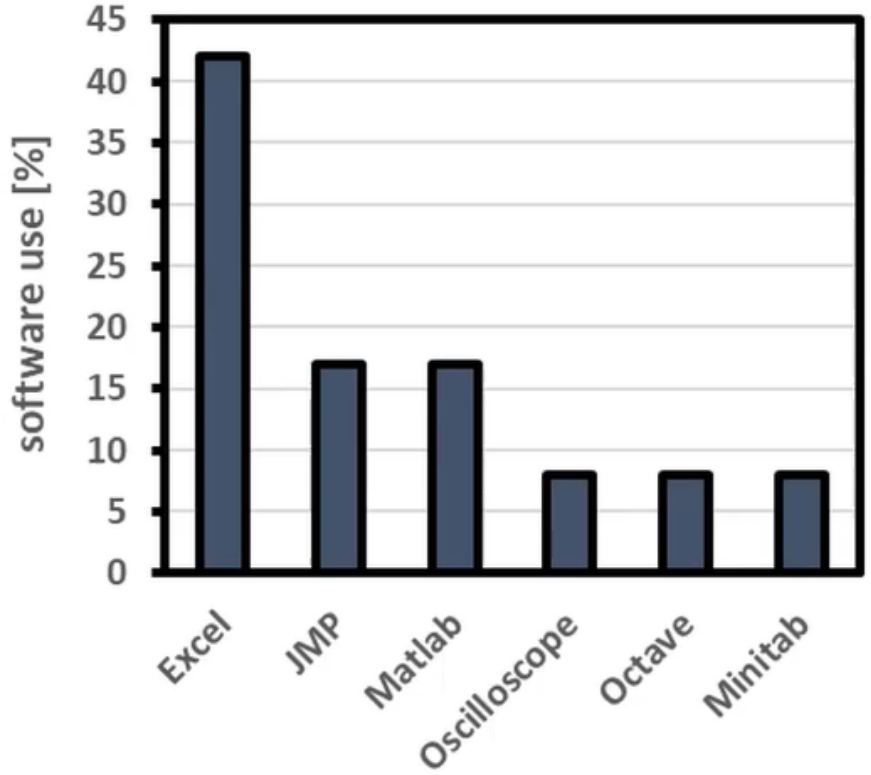
not appropriate



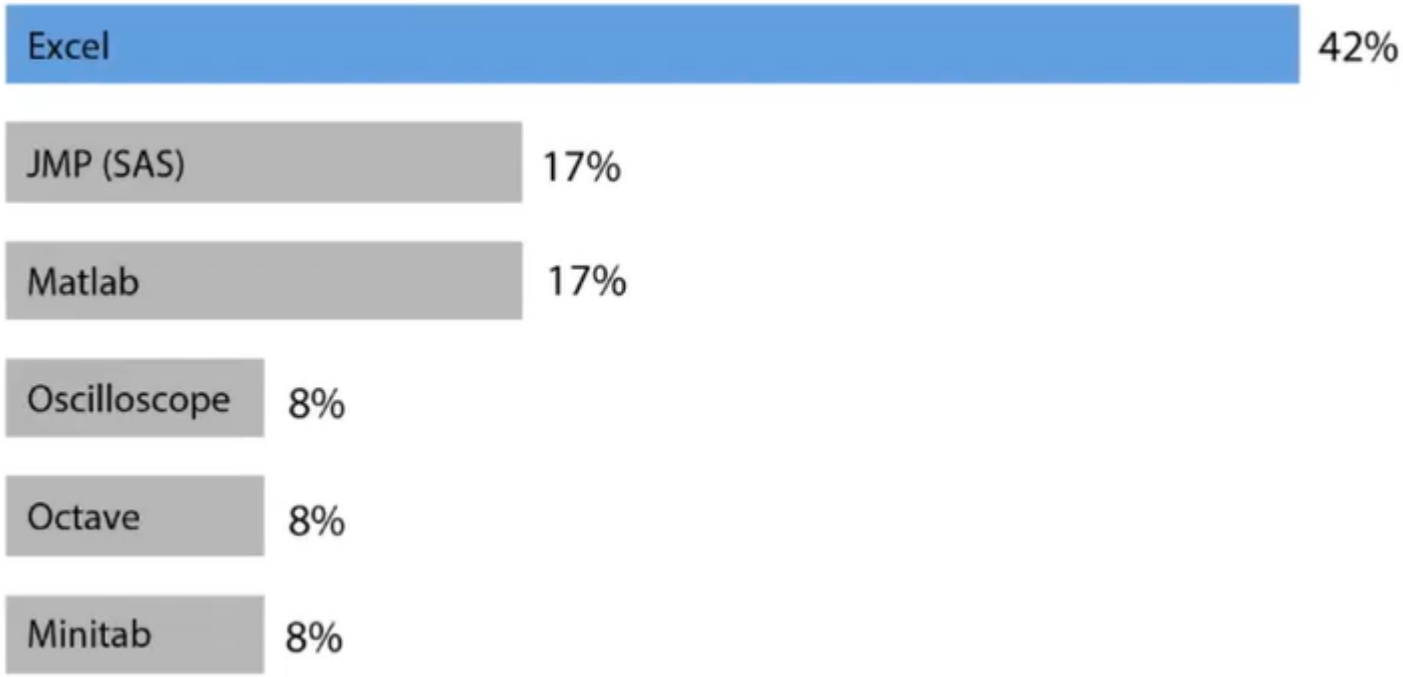
Inappropriate plots or tables

Another example

better



maybe the best



Inappropriate plots or tables

Here: the example to explain what properties are in a database

1

Soil temperature regimes						
Groups	**Frigid**	**Mesic**	**Thermic**	**Isothermic**	**Hyperthermic**	**Isohyperthe**
Number	4695	15240	8662	502	1455	3409
Soil texture classes						
Groups	**Sand**	**Loamy sand**	**Sandy loam**	**Loam**	**Silt loam**	**Sandy clay l**
Number	885	1552	5769	4758	5623	1843
Groups	**Clay loam**	**Silty clay loam**	**Silty clay**	**Clay**		
Number	2748	3230	2116	4157		
Soil depths (cm)						
Groups	**0-10**	**10-20**	**20-30**	**30-40**	**40-50**	**50-60**
Number	3876	3448	2915	2751	2289	2083
Groups	**60-70**	**70-80**	**80-90**	**90-100**	**100-110**	**110-120**
Number	1837	1684	1507	1389	1284	1172
Groups	**120-130**	**130-140**	**140-150**	**150-160**	**160-170**	**170-180**
Number	1105	1145	959	720	659	573

2

- Too many numbers
- Not easy to identify which soil class or depth is most frequent

- Maybe better to shift to the Supplementary

Inappropriate plots or tables

Here: the example to explain what properties are in a database

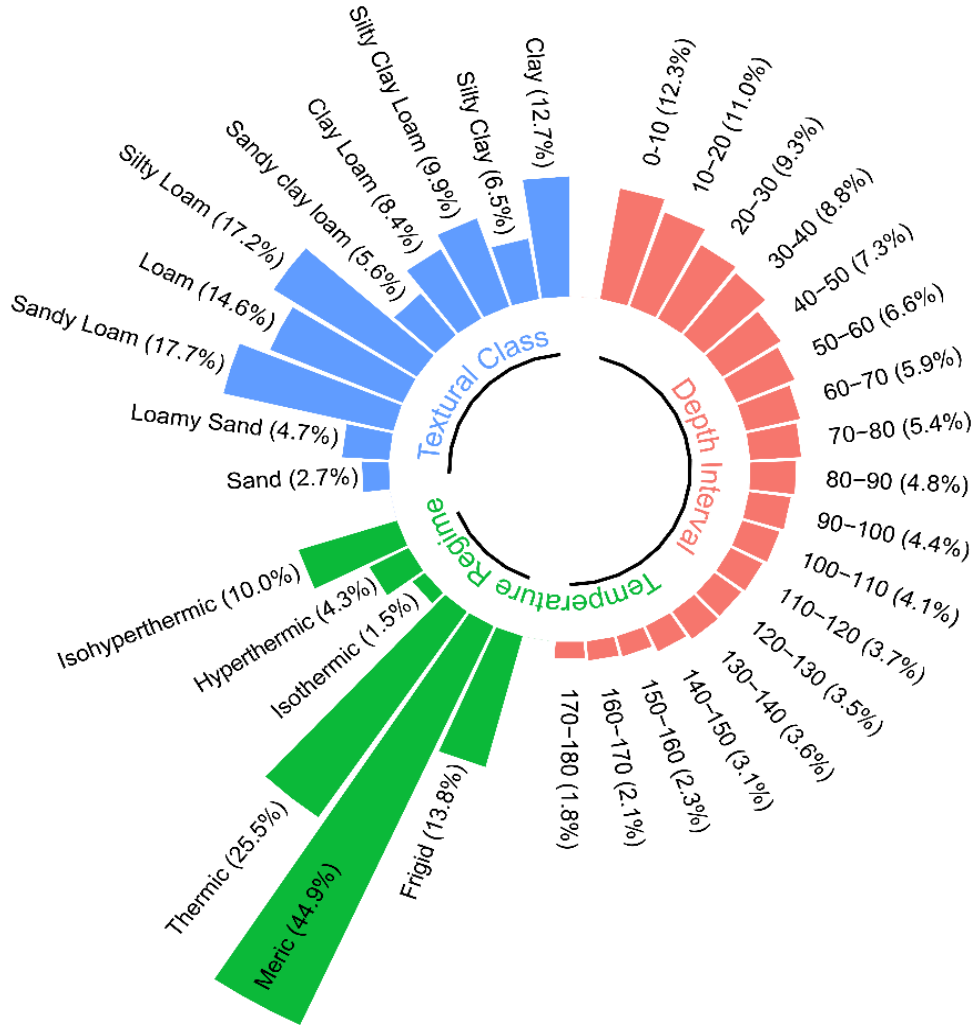
1

Soil temperature regimes						
Groups	**Frigid**	**Mesic**	**Thermic**	**Isothermic**	**Hyperthermic**	**Isohyperthe**
Number	4695	15240	8662	502	1455	3409

Soil texture classes						
Groups	**Sand**	**Loamy sand**	**Sandy loam**	**Loam**	**Silt loam**	**Sandy clay l**
Number	885	1552	5769	4758	5623	1843
Groups	**Clay loam**	**Silty clay loam**	**Silty clay**	**Clay**		
Number	2748	3230	2116	4157		

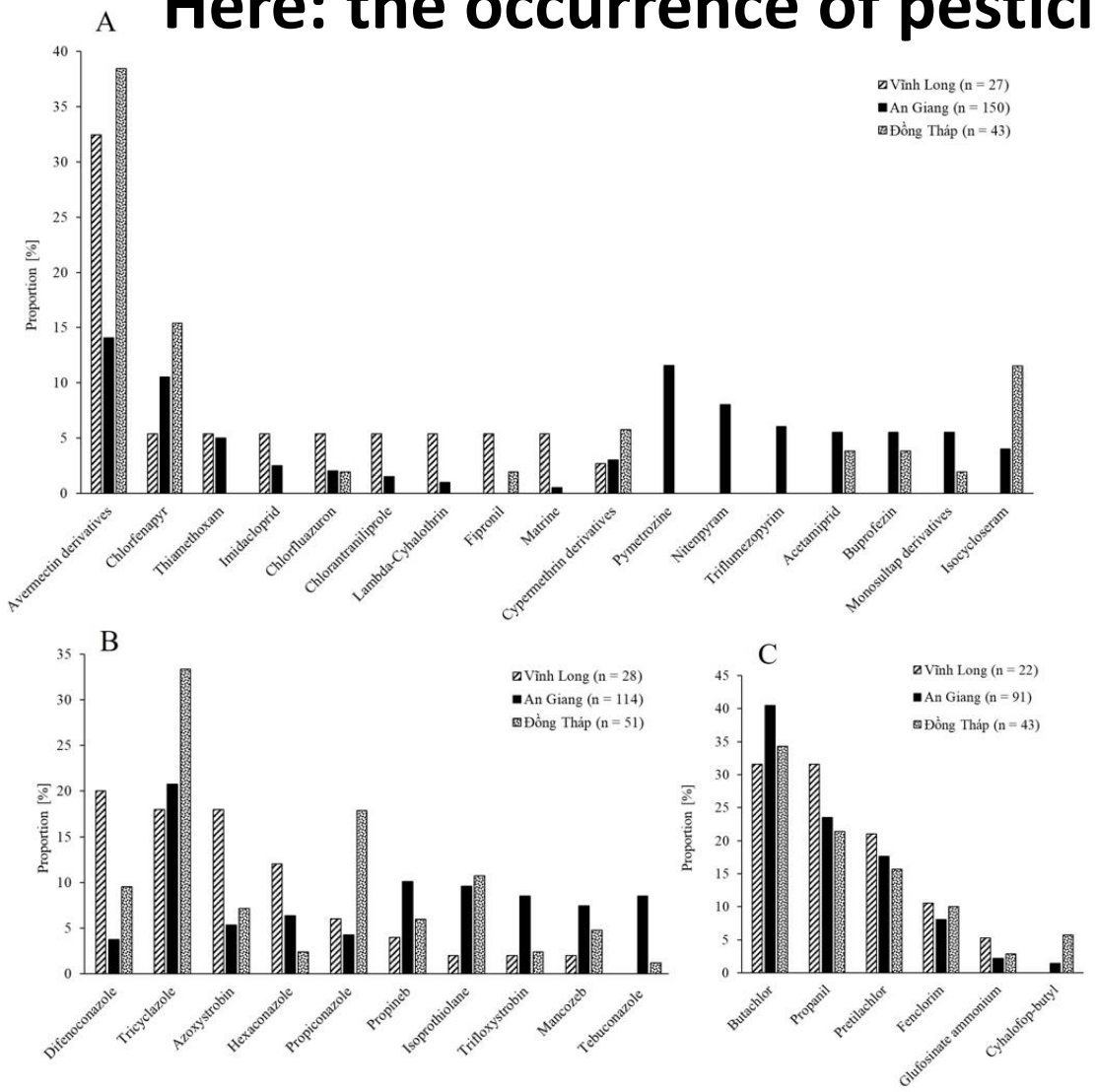
Soil depths (cm)						
Groups	**0-10**	**10-20**	**20-30**	**30-40**	**40-50**	**50-60**
Number	3876	3448	2915	2751	2289	2083
Groups	**60-70**	**70-80**	**80-90**	**90-100**	**100-110**	**110-120**
Number	1837	1684	1507	1389	1284	1172
Groups	**120-130**	**130-140**	**140-150**	**150-160**	**160-170**	**170-180**
Number	1105	1145	959	720	659	573

2



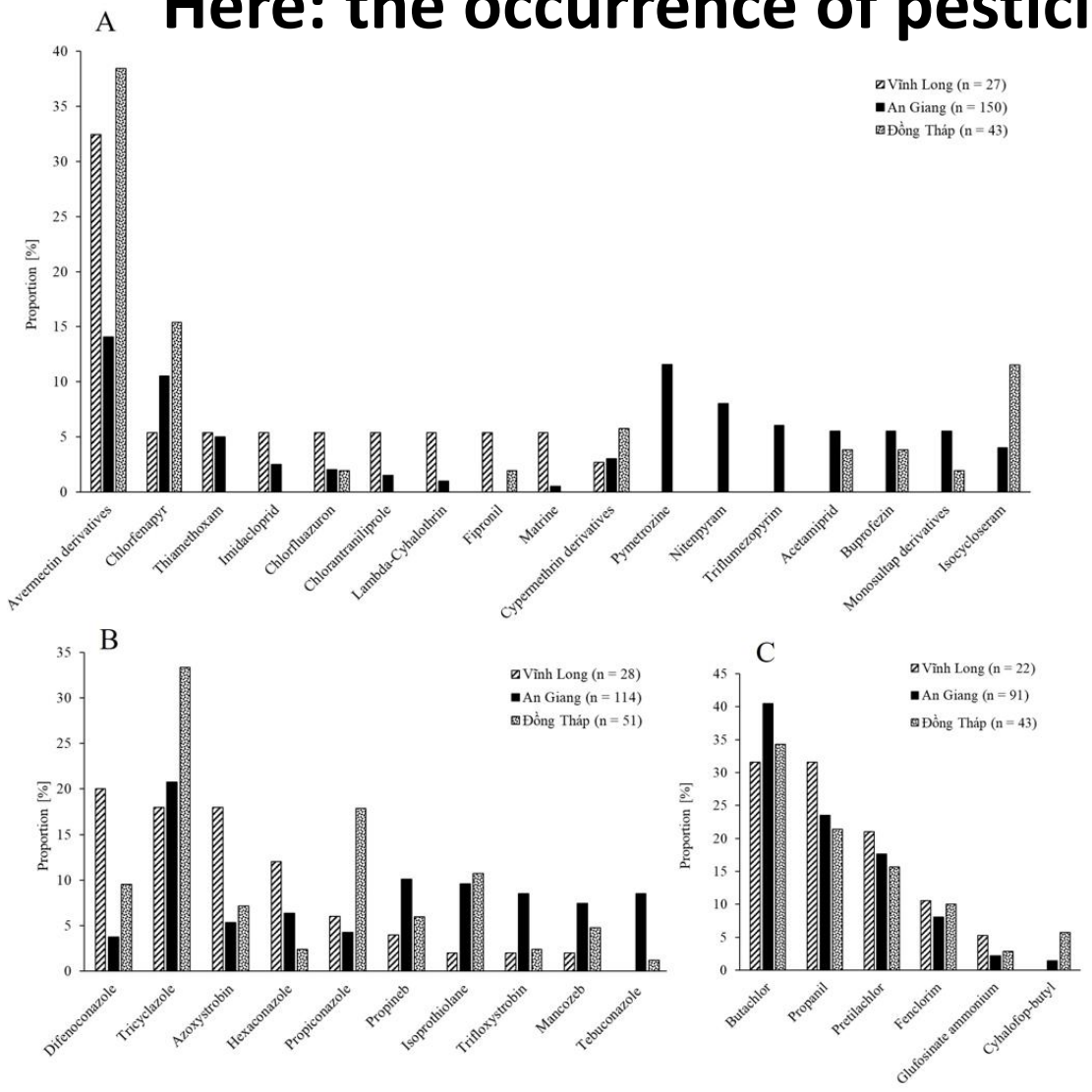
Inappropriate plots or tables

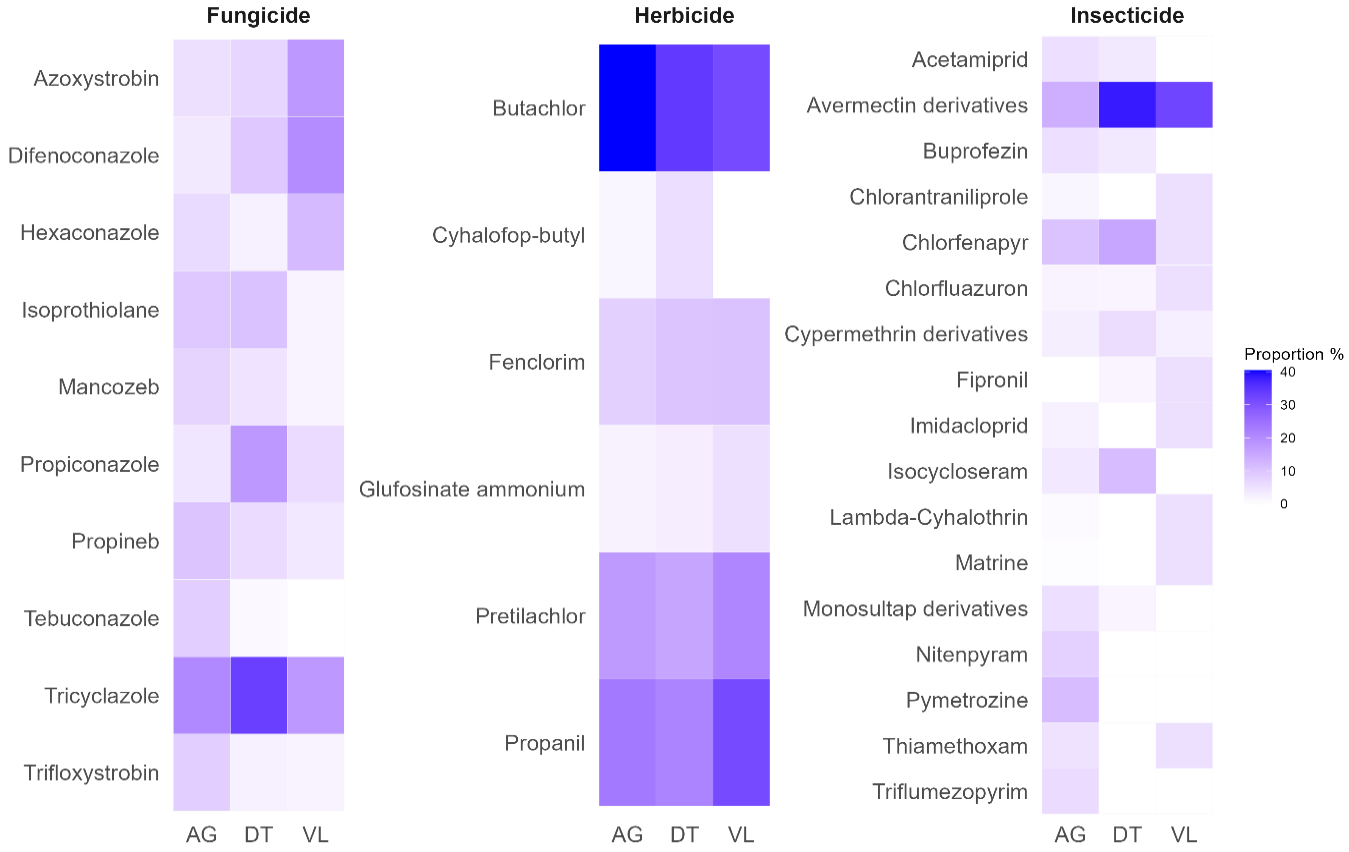
Here: the occurrence of pesticides in different provinces



Inappropriate plots or tables

Here: the occurrence of pesticides in different provinces

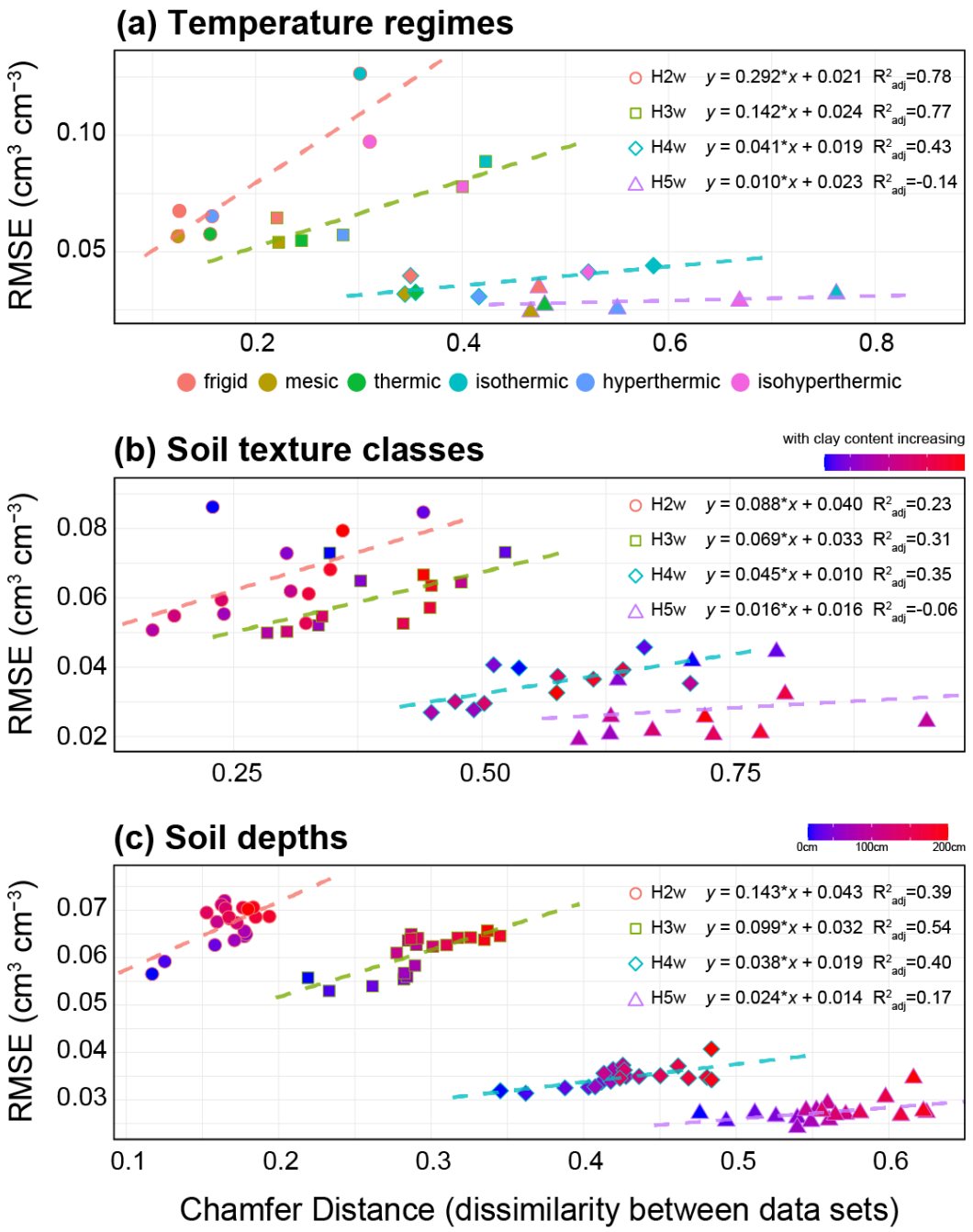




Plots should be self explainable

Therefore, the caption needs to provide all necessary information

Figure X: Chamfer Distance (CD) vs. RMSE between PTF estimated and measured soil water contents for reported pressure heads taken from the NCSS data set. (a) temperature regimes with six subsets from frigid to isohyperthermic, (b) 12 USDA soil texture classes, and (c) 20 soil depths were used to subdivide the NCSS data set. Points are the different shapes indicate the different Rosetta3 PTFs (H2w: cricle, H3w: square, H4w: rhombus, H5w: triangle), and colors indicate the different subsets. Fitted curves with different colors are shown for each PTF with corresponding R^2_{adj} and regression equation.



Don’t be boring

Now Lutz takes over

Don’t be boring

Stefens et al (2019) showed in their study……. Wang et al. (1995) analysed in their study……. Than (1985) showed in their study……

The study of Stefens et al (2019) showed …… and the results of Wang et al. (1995) support the same pattern. On the other hand, Than (1985) draw a different conclusion…..

That means that you have to combine the individual statements and line out a new narrative

Don’t be boring

Do not explain all details (numbers) you see in the table

Field	Sensors data used	Training (70% of data used)			Validation (30% of data used)		
		R^2	RMSE	NSE	R^2	RMSE	NSE
LT_RUM	UAV	0.938	0.106	0.934	0.837	0.039	0.791
	EMI	0.807	0.200	0.787	0.366	0.347	0.238
	S2	0.952	0.098	0.946	0.858	0.146	0.857
	UAV + EMI	0.922	0.123	0.918	0.511	0.295	0.467
	UAV + S2	0.961	0.086	0.951	0.659	0.224	0.630
	EMI + S2	0.971	0.067	0.988	0.605	0.240	0.579
	UAV + EMI + S2	0.942	0.087	0.978	0.667	0.222	0.639
LT_VIE	UAV	0.930	0.094	0.851	0.411	0.208	0.256
	EMI	0.756	0.129	0.714	0.191	0.251	0.097
	S2	0.902	0.100	0.840	0.149	0.189	0.130
	UAV + EMI	0.879	0.099	0.833	0.270	0.209	0.243
	UAV + S2	0.936	0.109	0.809	0.146	0.209	0.065
	EMI + S2	0.908	0.106	0.819	0.225	0.179	0.219
	UAV + EMI + S2	0.932	0.106	0.821	0.210	0.206	0.035
LT_VEZ	UAV	0.900	0.079	0.880	0.543	0.149	0.527
	EMI	0.779	0.113	0.756	0.149	0.219	0.068
	S2	0.957	0.081	0.875	0.196	0.27	0.263
	UAV + EMI	0.907	0.087	0.863	0.595	0.133	0.588
	UAV + S2	0.942	0.089	0.851	0.728	0.188	0.390
	EMI + S2	0.940	0.071	0.904	0.011	0.278	0.138
	UAV + EMI + S2	0.947	0.066	0.918	0.175	0.224	0.132

A student of mine described the data of the table over more than 1 page (and had 5 of those tables)

Don’t be boring

Do not explain all details (numbers) you see in the table

Field	Sensors data used	Training (70% of data used)			Validation (30% of data used)		
		R^2	RMSE	NSE	R^2	RMSE	NSE
LT_RUM	UAV	0.938	0.106	0.934	0.837	0.039	0.791
	EMI	0.807	0.200	0.787	0.3		38
	S2	0.952	0.098	0.946	0.858	14	0.857
	UAV + EMI	0.922	0.123	0.918	1	0.295	0.467
	UAV + S2	0.961	0.086	0.951	6	0.224	0.630
	EMI + S2	0.971	0.067	8	605	0.240	0.579
	UAV + EMI + S2	0.942	0.087	0.9	0.667	0.222	0.639
LT_VIE	UAV	0.930	094	0.851	0.411	0.208	0.256
	EMI	0.756	9	0.714	0.191	0.251	0.097
	S2	0.902	10	840	0.149	0.189	0.130
	UAV + EMI	9	099	0.833	0.270	0.209	0.243
	UAV + S2	93	0.109	0.809	0.146	0.209	0.065
	EMI + S	8	0.106	0.819	0.225	0.179	0.219
	UAV + EMI + S2	0.932	0.106	0.821	0.210	0.206	0.035
VEZ	UAV	0.900	0.079	0.880	0.543	0.149	0.527
	EMI	0.779	0.113	0.756	0.149	0.219	0.068
	S2	0.957	0.081	0.875	0.196	0.27	0.263
	UAV + EMI	0.907	0.087	0.863	0.595	0.133	0.588
	UAV + S2	0.942	0.089	0.851	0.728	0.188	0.390
	EMI + S2	0.940	0.071	0.904	0.011	0.278	0.138
	UAV + EMI + S2	0.947	0.066	0.918	0.175	0.224	0.132

Extremely boring

A student of mine described the data of the table over more than 1 page (and had 5 of those tables)

Don’t be boring

Discuss highest and lowest numbers (and provide these numbers in text)

Field	Sensors data used	Training (70% of data used)			Validation (30% of data used)		
		R^2	RMSE	NSE	R^2	RMSE	NSE
LT_RUM	UAV	0.938	0.106	0.934	0.837	0.039	0.791
	EMI	0.807	0.200	0.787	0.366	0.347	0.238
	S2	0.952	0.098	0.946	0.858	0.146	0.857
	UAV + EMI	0.922	0.123	0.918	0.511	0.295	0.467
	UAV + S2	0.961	0.086	0.951	0.659	0.224	0.630
	EMI + S2	0.971	0.067	0.988	0.605	0.240	0.579
	UAV + EMI + S2	0.942	0.087	0.978	0.667	0.222	0.639
LT_VIE	UAV	0.930	0.094	0.851	0.411	0.208	0.256
	EMI	0.756	0.129	0.714	0.191	0.251	0.097
	S2	0.932	0.100	0.840	0.149	0.189	0.130
	UAV + EMI	0.879	0.099	0.833	0.270	0.209	0.243
	UAV + S2	0.936	0.109	0.809	0.146	0.209	0.065
	EMI + S2	0.908	0.106	0.819	0.225	0.179	0.219
	UAV + EMI + S2	0.932	0.106	0.821	0.210	0.206	0.035
LT_VEZ	UAV	0.900	0.079	0.880	0.543	0.149	0.527
	EMI	0.779	0.113	0.756	0.149	0.219	0.068
	S2	0.957	0.081	0.875	0.196	0.27	0.263
	UAV + EMI	0.907	0.087	0.863	0.595	0.133	0.588
	UAV + S2	0.942	0.089	0.851	0.728	0.188	0.390
	EMI + S2	0.940	0.071	0.904	0.011	0.278	0.138
	UAV + EMI + S2	0.947	0.066	0.918	0.479	0.224	0.132

Don’t be boring

Discuss differences (high consistency versus low consistency)

Field	Sensors data used	Training (70% of data used)			Validation (30% of data used)		
		R²	RMSE	NSE	R²	RMSE	NSE
LT_RUM	UAV	0.938	0.106	0.934	0.837	0.039	0.791
	EMI	0.807	0.200	0.787	0.366	0.347	0.238
	S2	0.952	0.098	0.946	0.858	0.146	0.857
	UAV + EMI	0.922	0.123	0.918	0.511	0.295	0.467
	UAV + S2	0.961	0.086	0.951	0.659	0.224	0.630
	EMI + S2	0.971	0.067	0.988	0.605	0.240	0.579
	UAV + EMI + S2	0.94	0.087	0.978	0.66	0.222	0.639
LT_VIE	UAV	0.930	0.094	0.851	0.411	0.208	0.256
	EMI	0.756	0.129	0.714	0.191	0.251	0.097
	S2	0.902	0.100	0.840	0.149	0.189	0.130
	UAV + EMI	0.879	0.099	0.833	0.270	0.209	0.243
	UAV + S2	0.936	0.109	0.809	0.146	0.209	0.065
	EMI + S2	0.908	0.106	0.819	0.225	0.179	0.219
	UAV + EMI + S2	0.937	0.106	0.821	0.210	0.206	0.035
LT_VEZ	UAV	0.900	0.079	0.880	0.543	0.149	0.527
	EMI	0.779	0.113	0.756	0.149	0.219	0.068
	S2	0.957	0.081	0.875	0.196	0.27	0.263
	UAV + EMI	0.907	0.087	0.863	0.595	0.133	0.588
	UAV + S2	0.942	0.089	0.851	0.728	0.188	0.390
	EMI + S2	0.940	0.071	0.904	0.011	0.278	0.138
	UAV + EMI + S2	0.947	0.066	0.918	0.175	0.224	0.132

and discuss why this happens (reasons) or if this is in-line with other findings

Misuse of abbreviations

Abbreviations should be used sparingly. Although abbreviations make the writing more concise, they also make the writing less clear, because a reader may not remember what an abbreviation means, particularly when excessive numbers of abbreviations are used.

Abbreviation should be used to shorten long words or word combinations that are used repeatedly in an article. **Short words should not be abbreviated**. And, abbreviations should, at the very least, consist of letter that actually occur in the word that is abbreviated.

Misuse of abbreviations

not appropriate

VN	**= Vinh Long**	**-> as Vinh Long is already short. I also recommend not to abbreviate locations as it sometimes confuses the readers**
Crtl	**= control**	**-> not much shorter as full word**
CK	**= control**	**-> often used in literature but control does not contain a K)**
bS	**= bulk sulfur**	**-> not a standard abbreviation**
iS	**= ionic sulfate**	**-> some here: Could be also ionic strength**

…….

Misuse of abbreviations

appropriate

FC = field capacity
PAW = plant available water
AGB = above ground biomass
NEE = net ecosystem exchange
ET = evapotranspiration
SWRC = soil water retention curve
RS = remote sensing
AI = artificial intelligence
DAS = days after sowing
SOC = soil organic carbon
TOC = total organic carbon
C_{org} = organic carbon
SWC = soil water content
……

and abbreviations are topic based

Remote sensing AI = Artificial Intelligence

Pesticide research ai = active ingredient

Misuse of abbreviations

appropriate

FC = field capacity
PAW = plant available water
AGB = above ground biomass
NEE = net ecosystem exchange
ET = evapotranspiration
SWRC = soil water retention curve
RS = remote sensing
AI = artificial intelligence
DAS = days after sowing
SOC = soil organic carbon
TOC = total organic carbon
C_{org} = organic carbon
SWC = soil water content
……

and abbreviations are topic based

Remote sensing AI = Artificial Intelligence

Pesticide Research ai = active ingredient

In any case all abbreviations have to be introduced

Start a writing club

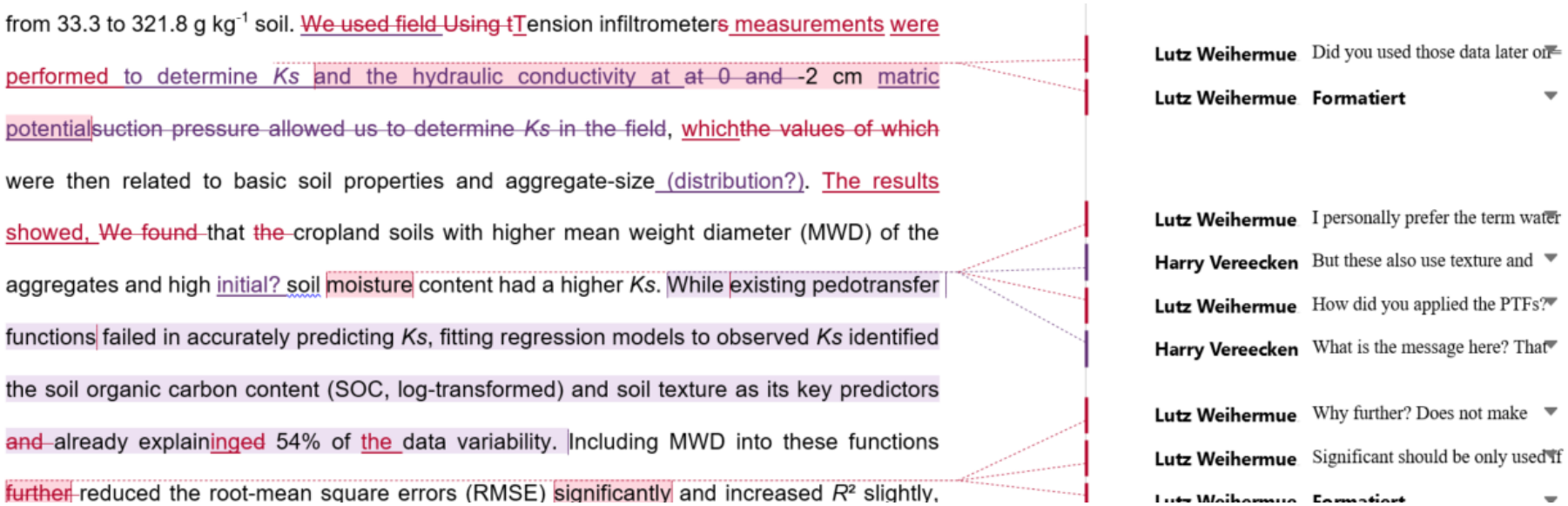
Share early version of your manuscript (or even parts such as M&M) with colleagues for internal review and discussion

Revise manuscripts you receive rigorously and comment wherever you do not understand

Check manuscript for consistency and easy narrative

Be polite during the internal review but do not hesitate to criticise /correct

Start a writing club



And do not be disappointed if your text looks like this

Finally, start up writing