Molecular Biotechnology and Bioinformatics

สถานวิจัยจีโนมและชีวสารสนเทศ มหาวิทยาลัยสงขลานครินทร์

Agricultural Biotech & Bioinformatics Lab

Omics Data Handling and Visualization in R

17.5.2023
Sci.PSU

Agricultural Biotech & Bioinformatics Lab,
Division of Biological Science, Faculty of Sciences PSU

Agricultural Biotech & Bioinformatics Lab, PSU

SCIENCE

https://molbioshrimp.wordpress.com/

ponsit.sat@gmail.com, ponsit.s@psu.ac.th

Molecular Biotechnology and Bioinformatics

สถานวิจัยจีโนมและชีวสารสนเทศ
มหาวิทยาลัยสงขลานครินทร์

Agricultural Biotech & Bioinformatics Lab

Speakers and teaching assistants

Asst. Prof. Ponsit Sathapondecha, PhD

Jiratchaya Nuanpirom, MSc

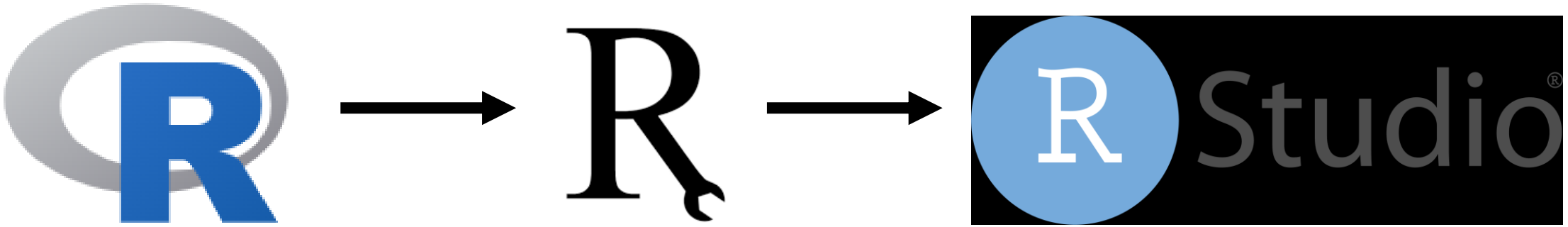
Prasert Yodsawat, MSc

Khunanon Chanasongkhram, BSc

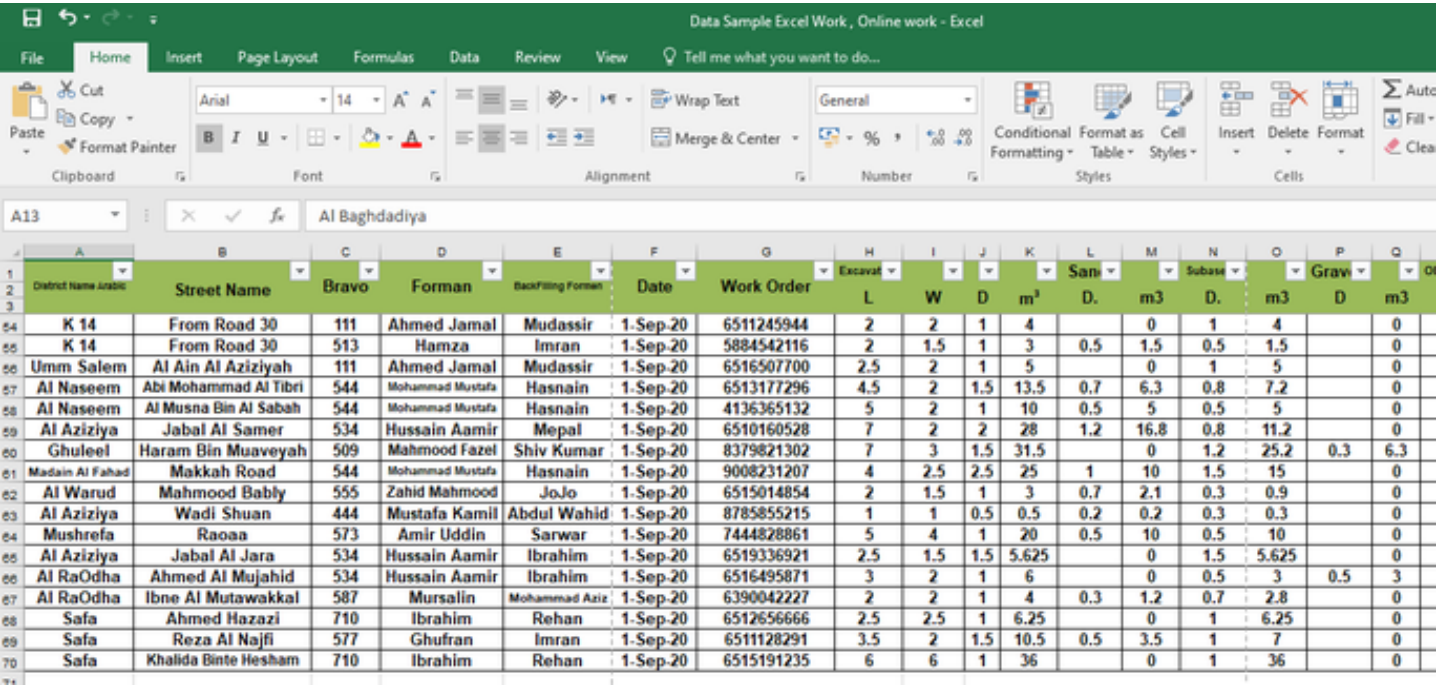
PSU SCIENCE

Online Handout & R installation

- https://jirathnuan.github.io/r-handviz-workshop/schedule.html
- https://jirathnuan.github.io/r-handviz-workshop/prerequisite.html



Introduction to Data Visualization



District Name Arabic	Street Name	Bravo	Forman	BackFilling Forman	Date	Work Order	Excavat L	W	D	m³	San D.	m3	Subase D.	m3	Grav D	m3
K 14	From Road 30	111	Ahmed Jamal	Mudassir	1-Sep-20	6511245944	2	2	1	4		0	1	4		0
K 14	From Road 30	513	Hamza	Imran	1-Sep-20	5884542116	2	1.5	1	3	0.5	1.5	0.5	1.5		0
Umm Salem	Al Ain Al Aziziyah	111	Ahmed Jamal	Mudassir	1-Sep-20	6516507700	2.5	2	1	5		0	1	5		0
Al Naseem	Abi Mohammad Al Tibri	544	Mohammad Mustafa	Hasnain	1-Sep-20	6513177296	4.5	2	1.5	13.5	0.7	6.3	0.8	7.2		0
Al Naseem	Al Musna Bin Al Sabah	544	Mohammad Mustafa	Hasnain	1-Sep-20	4136365132	5	2	1	10	0.5	5	0.5	5		0
Al Aziziya	Jabal Al Samer	534	Hussain Aamir	Mepal	1-Sep-20	6510160528	7	2	2	28	1.2	16.8	0.8	11.2		0
Ghuleel	Haram Bin Muaveyah	509	Mahmood Fazel	Shiv Kumar	1-Sep-20	8379821302	7	3	1.5	31.5		0	1.2	25.2	0.3	6.3
Madain Al Fahad	Makkah Road	544	Mohammad Mustafa	Hasnain	1-Sep-20	9008231207	4	2.5	2.5	25	1	10	1.5	15		0
Al Warud	Mahmood Bably	555	Zahid Mahmood	JoJo	1-Sep-20	6515014854	2	1.5	1	3	0.7	2.1	0.3	0.9		0
Al Aziziya	Wadi Shuan	444	Mustafa Kamil	Abdul Wahid	1-Sep-20	8785855215	1	1	0.5	0.5	0.2	0.2	0.3	0.3		0
Mushrefa	Raoaa	573	Amir Uddin	Sarwar	1-Sep-20	7444828861	5	4	1	20	0.5	10	0.5	10		0
Al Aziziya	Jabal Al Jara	534	Hussain Aamir	Ibrahim	1-Sep-20	6519336921	2.5	1.5	1.5	5.625		0	1.5	5.625		0
Al RaOdha	Ahmed Al Mujahid	534	Hussain Aamir	Ibrahim	1-Sep-20	6516495871	3	2	1	6		0	0.5	3	0.5	3
Al RaOdha	Ibne Al Mutawakkal	587	Mursalin	Mohammad Aziz	1-Sep-20	6390042227	2	2	1	4	0.3	1.2	0.7	2.8		0
Safa	Ahmed Hazazi	710	Ibrahim	Rehan	1-Sep-20	6512656666	2.5	2.5	1	6.25		0	1	6.25		0
Safa	Reza Al Najfi	577	Ghufran	Imran	1-Sep-20	6511128291	3.5	2	1.5	10.5	0.5	3.5	1	7		0
Safa	Khalida Binte Hesham	710	Ibrahim	Rehan	1-Sep-20	6515191235	6	6	1	36		0	1	36		0

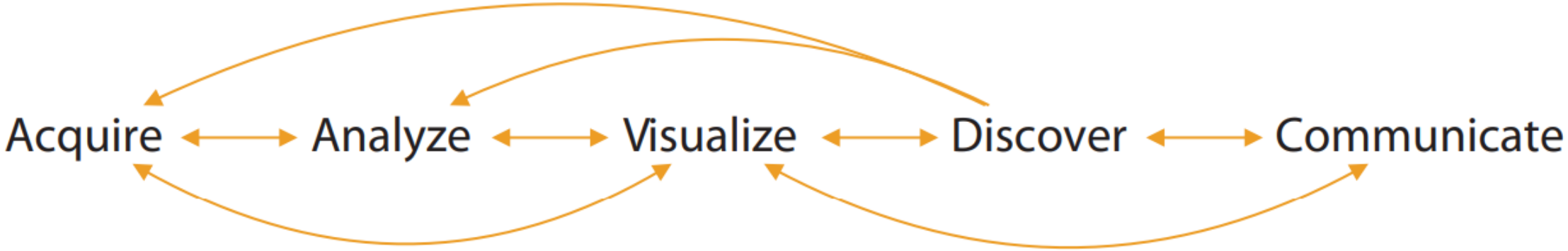
Data in excel

Presentation

Introduction to Data Visualization

Role of data visualization in research

Graphical representation of information and data.



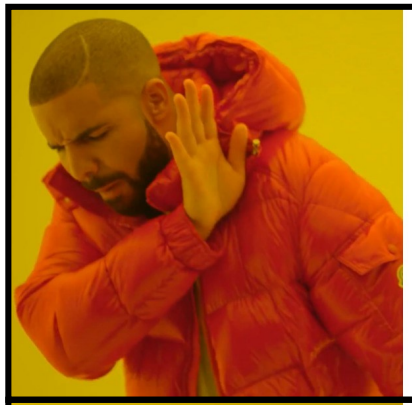
- Reveals pattern/trend of the data
- Allowing discover new insights into the data
- Facilitating communication

O'Donoghue, Seán I., et al. "Visualization of biomedical data." Annual Review of Biomedical Data Science 1 (2018): 275-304.

Introduction to Data Visualization

Misconceptions about data visualization

1 “The goal of data visualization is to impress”



- Think of data visualization is about ART only.
- Adding an optional wow factors not present in the data itself.
- Using gaudy colors than necessary.

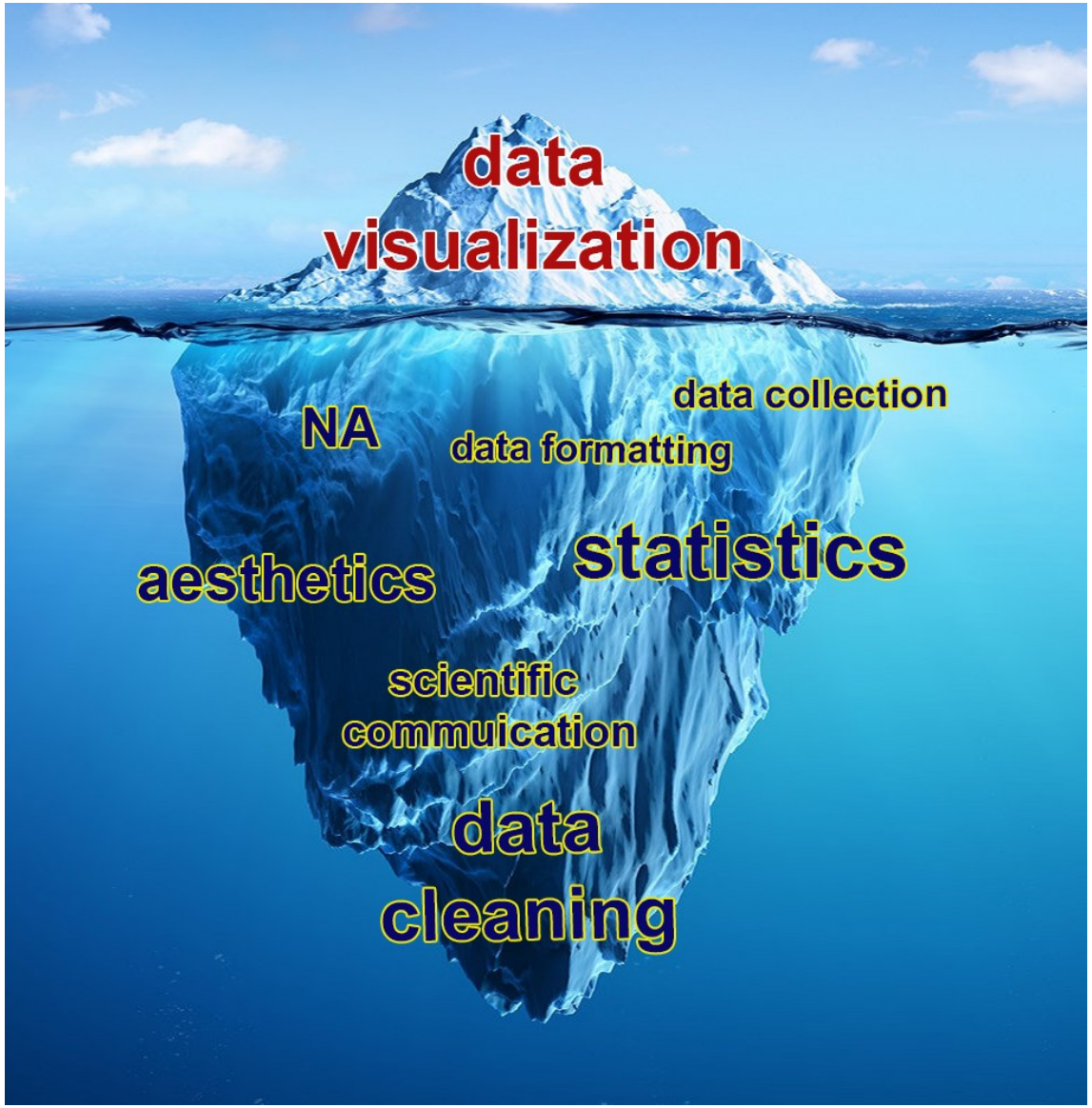
- Data visualization is about **ART and SCIENCE**, in order to make it glam and interpretable.
- The goal of data visualization is to reveal **patterns in data**.
- It would prefer to use **thematic or minimal colors**.

O'Donoghue, Seán I., et al. "Visualization of biomedical data." Annual Review of Biomedical Data Science 1 (2018): 275-304.

Introduction to Data Visualization

Misconceptions about data visualization

2 “Data visualization is easy”

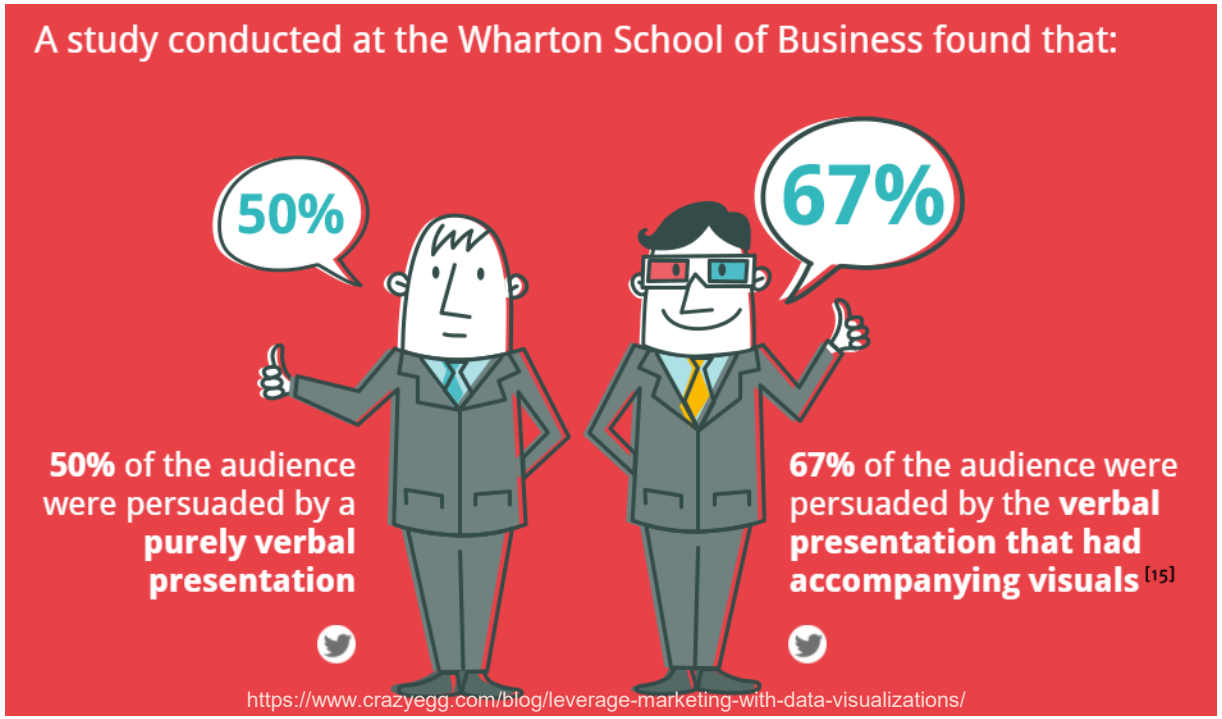


O'Donoghue, Seán I., et al. "Visualization of biomedical data." Annual Review of Biomedical Data Science 1 (2018): 275-304.

Introduction to Data Visualization

Misconceptions about data visualization

3 "Studying data visualization is unnecessary"



Free online data viz courses:

coursera 7-days free

- Data Visualization with R by IBM
- Data Visualization in R with ggplot2 by John Hopkins University

HARVARD UNIVERSITY edX

- Data Science: Visualization (free for audit)

Free all time

STHDA

R-bloggers

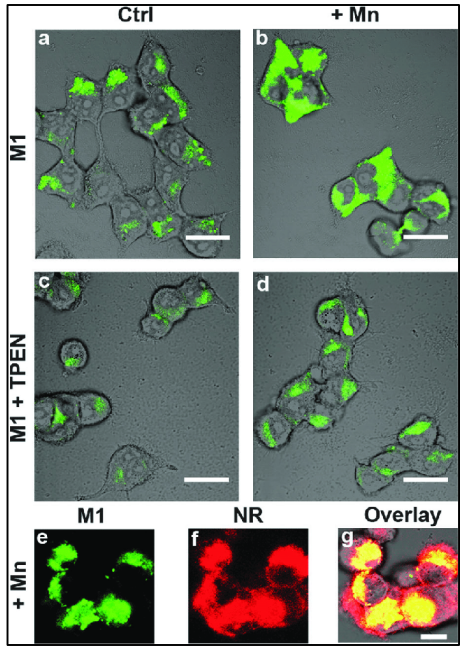
And many more!

O'Donoghue, Seán I., et al. "Visualization of biomedical data." Annual Review of Biomedical Data Science 1 (2018): 275-304.

Introduction to Data Visualization

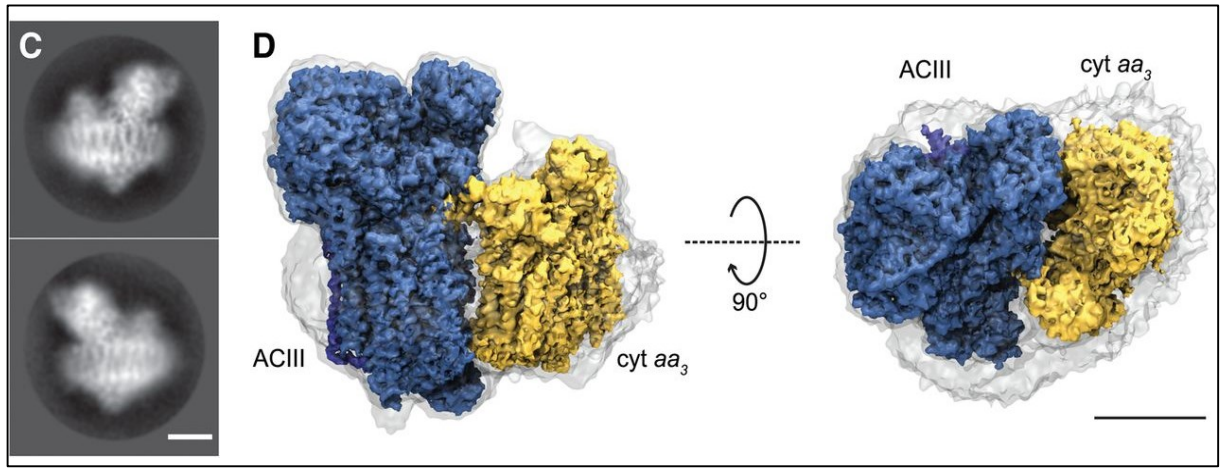
Misconceptions about data visualization

4 "VISUALIZATION == IMAGING"



PMID: 25572103

imaging

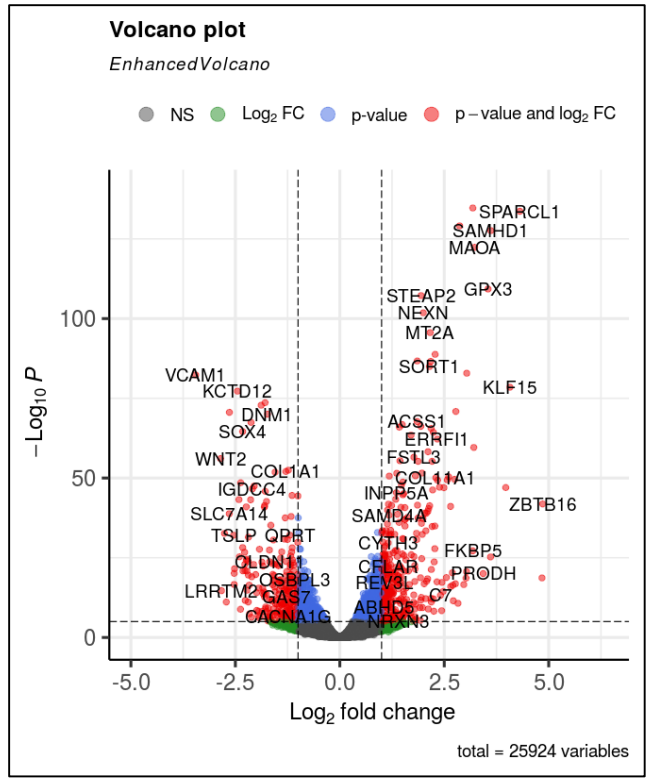


PMID: 30464048

visualization

imaging

visualization

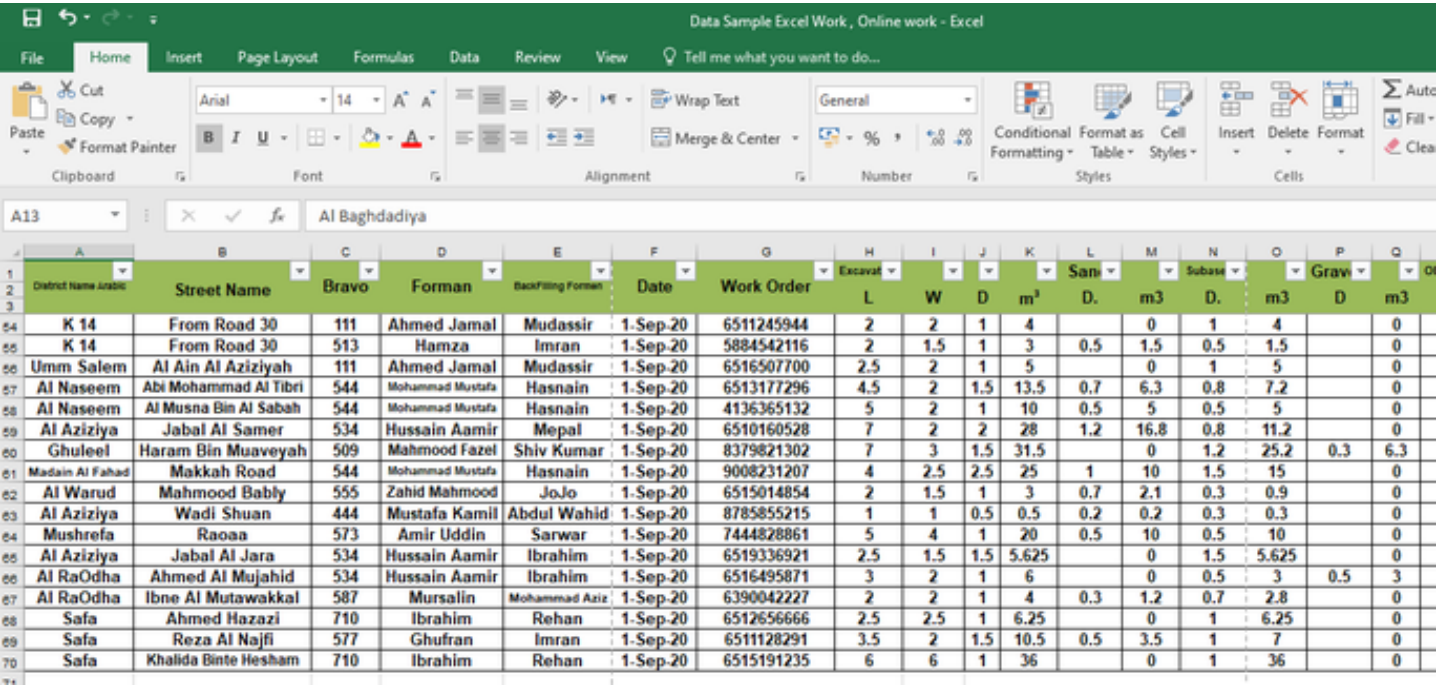


https://github.com/kevinblighe/EnhancedVolcano

visualization

O'Donoghue, Seán I., et al. "Visualization of biomedical data." Annual Review of Biomedical Data Science 1 (2018): 275-304.

Introduction to Data Visualization

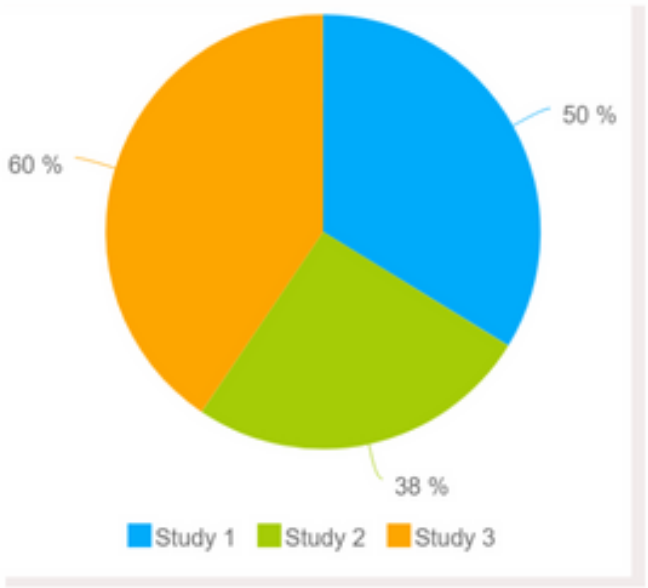


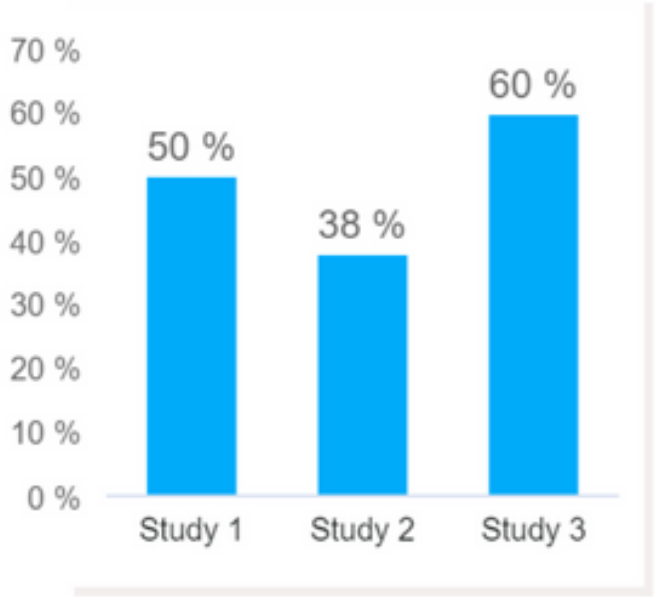
District Name Arabic	Street Name	Bravo	Forman	BackFilling Forman	Date	Work Order	Excavat L	W	D	m³	San D.	m3	Subase D.	m3	Grav D	m3
K 14	From Road 30	111	Ahmed Jamal	Mudassir	1-Sep-20	6511245944	2	2	1	4		0	1	4		0
K 14	From Road 30	513	Hamza	Imran	1-Sep-20	5884542116	2	1.5	1	3	0.5	1.5	0.5	1.5		0
Umm Salem	Al Ain Al Aziziyah	111	Ahmed Jamal	Mudassir	1-Sep-20	6516507700	2.5	2	1	5		0	1	5		0
Al Naseem	Abi Mohammad Al Tibri	544	Mohammad Mustafa	Hasnain	1-Sep-20	6513177296	4.5	2	1.5	13.5	0.7	6.3	0.8	7.2		0
Al Naseem	Al Musna Bin Al Sabah	544	Mohammad Mustafa	Hasnain	1-Sep-20	4136365132	5	2	1	10	0.5	5	0.5	5		0
Al Aziziya	Jabal Al Samer	534	Hussain Aamir	Mepal	1-Sep-20	6510160528	7	2	2	28	1.2	16.8	0.8	11.2		0
Ghuleel	Haram Bin Muaveyah	509	Mahmood Fazel	Shiv Kumar	1-Sep-20	8379821302	7	3	1.5	31.5		0	1.2	25.2	0.3	6.3
Madain Al Fahad	Makkah Road	544	Mohammad Mustafa	Hasnain	1-Sep-20	9008231207	4	2.5	2.5	25	1	10	1.5	15		0
Al Warud	Mahmood Bably	555	Zahid Mahmood	JoJo	1-Sep-20	6515014854	2	1.5	1	3	0.7	2.1	0.3	0.9		0
Al Aziziya	Wadi Shuan	444	Mustafa Kamil	Abdul Wahid	1-Sep-20	8785855215	1	1	0.5	0.5	0.2	0.2	0.3	0.3		0
Mushrefa	Raoaa	573	Amir Uddin	Sarwar	1-Sep-20	7444828861	5	4	1	20	0.5	10	0.5	10		0
Al Aziziya	Jabal Al Jara	534	Hussain Aamir	Ibrahim	1-Sep-20	6519336921	2.5	1.5	1.5	5.625		0	1.5	5.625		0
Al RaOdha	Ahmed Al Mujahid	534	Hussain Aamir	Ibrahim	1-Sep-20	6516495871	3	2	1	6		0	0.5	3	0.5	3
Al RaOdha	Ibne Al Mutawakkal	587	Mursalin	Mohammad Aziz	1-Sep-20	6390042227	2	2	1	4	0.3	1.2	0.7	2.8		0
Safa	Ahmed Hazazi	710	Ibrahim	Rehan	1-Sep-20	6512656666	2.5	2.5	1	6.25		0	1	6.25		0
Safa	Reza Al Najfi	577	Ghufran	Imran	1-Sep-20	6511128291	3.5	2	1.5	10.5	0.5	3.5	1	7		0
Safa	Khalida Binte Hesham	710	Ibrahim	Rehan	1-Sep-20	6515191235	6	6	1	36		0	1	36		0

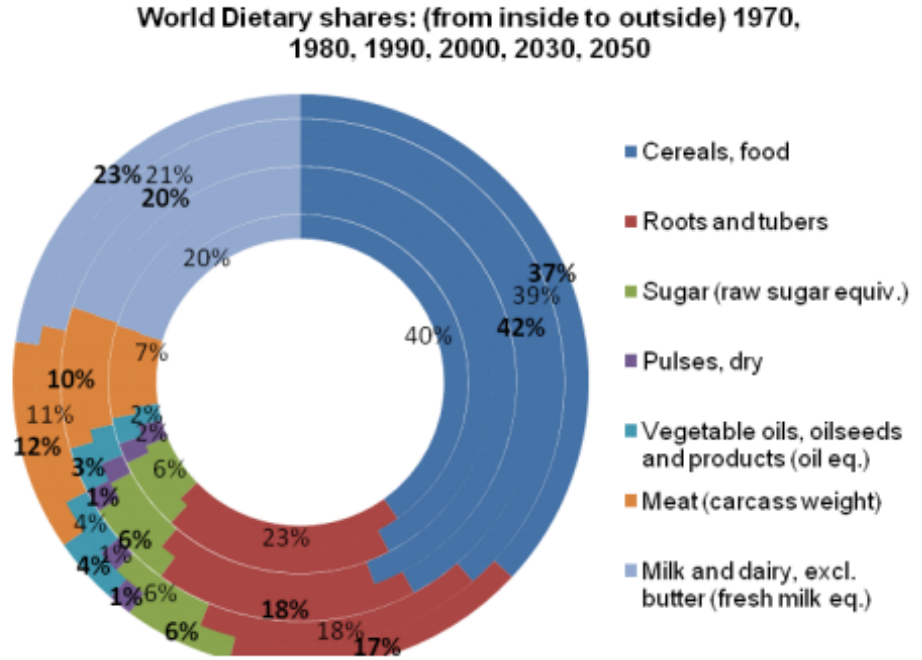
Data in excel

Visualization with excel

Choose a right chart



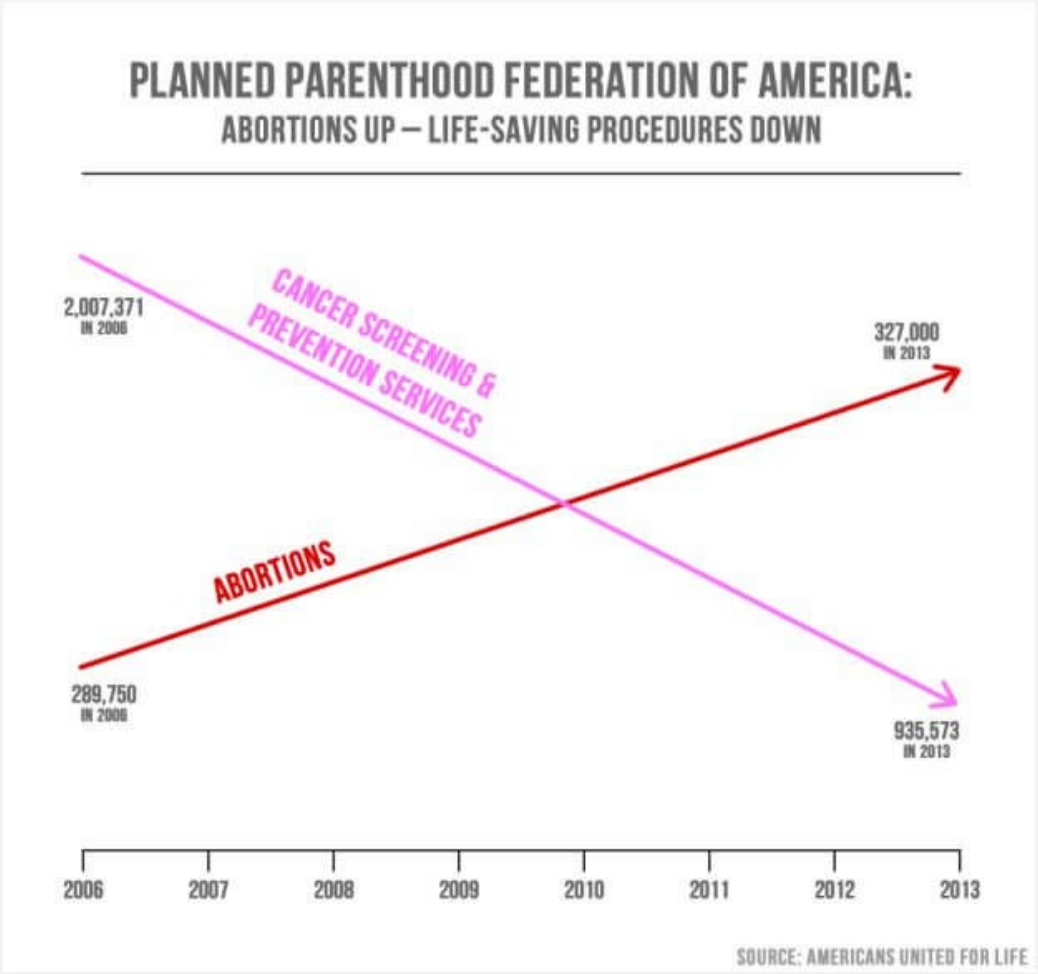




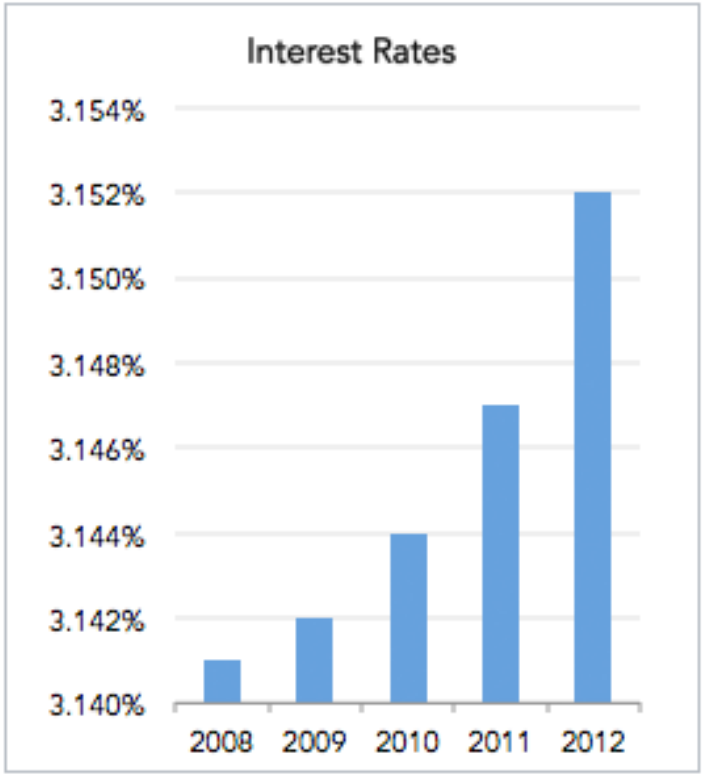
Note: figures for 1980 and 1990 shares are not shown for sake of clarity.

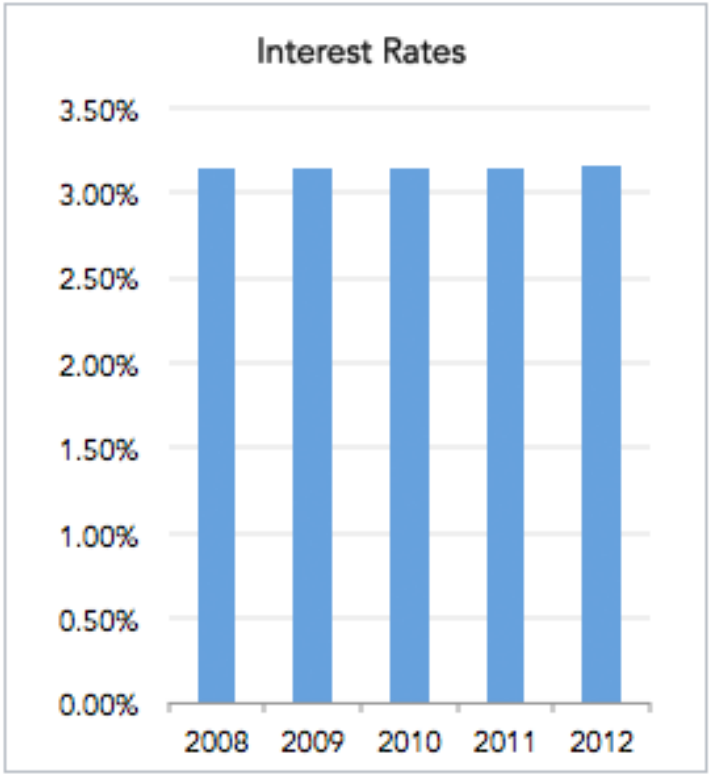
https://www.datapine.com/blog/misleading-data-visualization-examples/#how-to

Misleading consideration



Same Data, Different Y-Axis

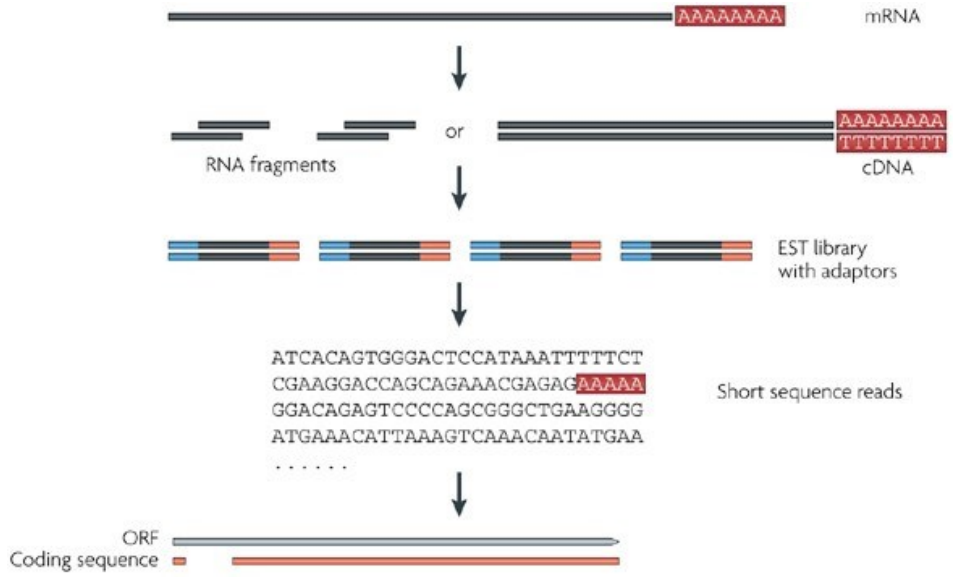


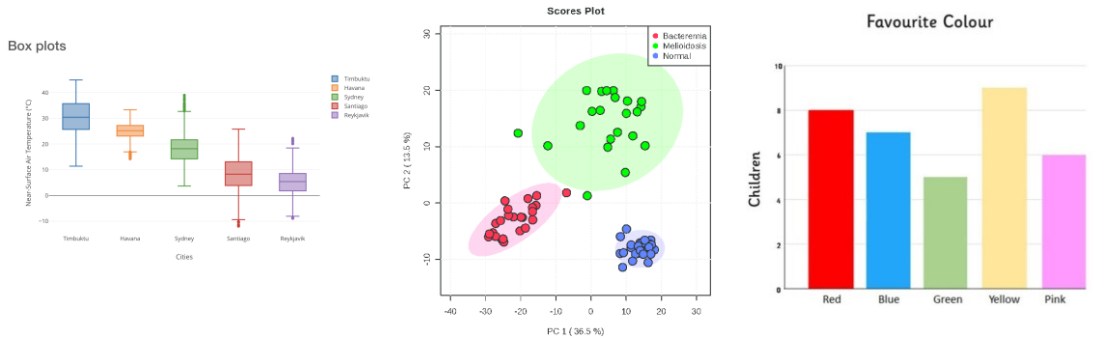


https://www.datapine.com/blog/misleading-data-visualization-examples/#how-to

How to obtain omics data

NGS: RNAseq, AmpliconSeq, Metgenomic-Seq, GBS etc.





QC data (intermediate data)

Post analyzed Data
Most common in matrix

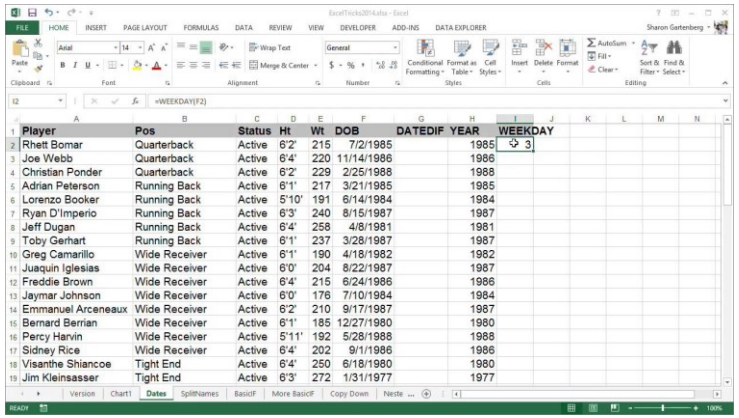
	Cell1	Cell2	...	CellN
Gene1	3	2	.	13
Gene2	2	3	.	1
Gene3	1	14	.	18
...	.	.	.	.
...	.	.	.	.
...	.	.	.	.
GeneM	25	0	.	0

?

Omics studies deals with big data!

Too big files for excel

Data manipulations

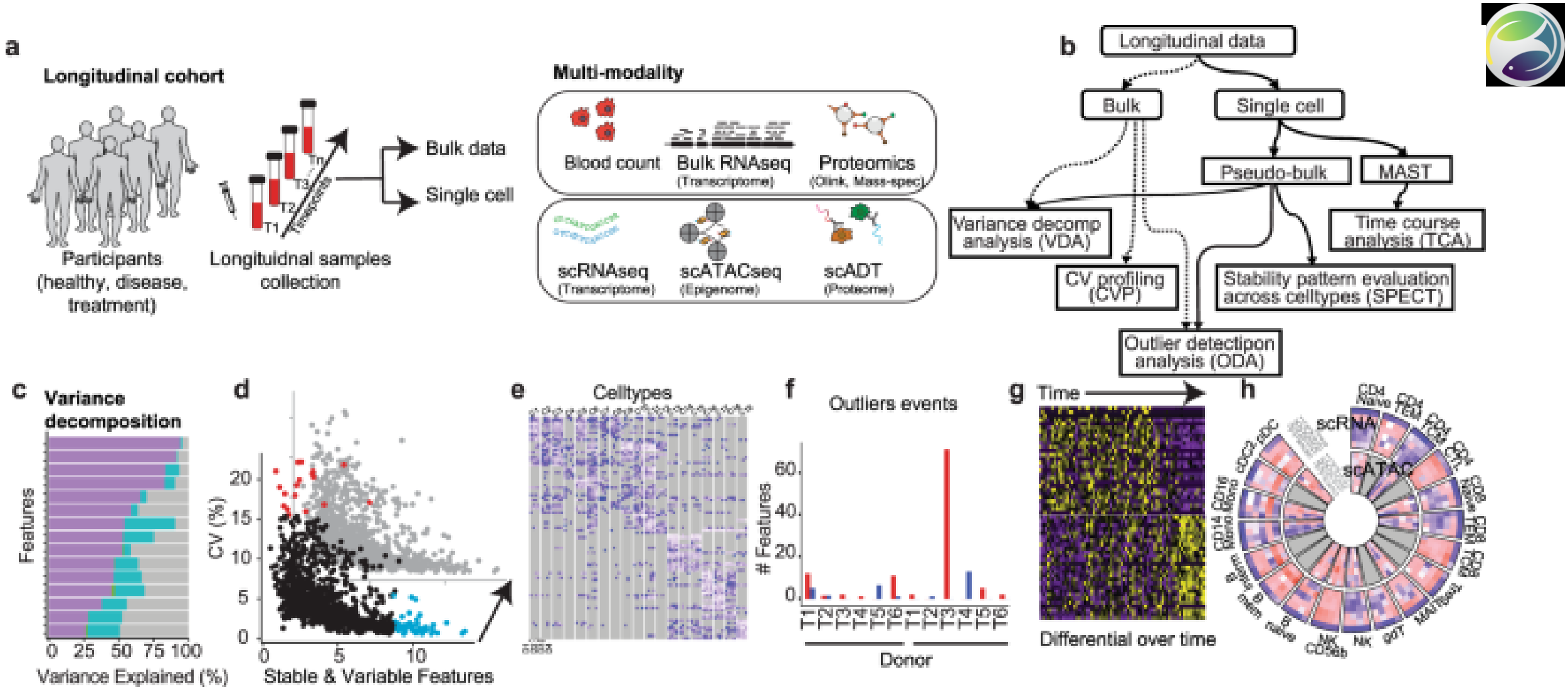


Player	Pos	Status	Ht	Wt	DOB	DATEDIF	YEAR	WEEKDAY
Rhett Bomar	Quarterback	Active	6'2'	215	7/2/1985		1985	3
Joe Webb	Quarterback	Active	6'4'	220	11/14/1986		1986	
Christian Ponder	Quarterback	Active	6'2'	229	2/25/1988		1988	
Adrian Peterson	Running Back	Active	6'1'	217	3/21/1985		1985	
Lorenzo Booker	Running Back	Active	5'10'	191	6/14/1984		1984	
Ryan D'Imperio	Running Back	Active	6'3'	240	8/15/1987		1987	
Jeff Dugan	Running Back	Active	6'4'	258	4/8/1981		1981	
Toby Gerhart	Running Back	Active	6'1'	237	3/28/1987		1987	
Greg Camarillo	Wide Receiver	Active	6'1'	190	4/18/1982		1982	
Juaquin Iglesias	Wide Receiver	Active	6'0'	204	8/22/1987		1987	
Freddie Brown	Wide Receiver	Active	6'4'	215	6/24/1986		1986	
Jaymar Johnson	Wide Receiver	Active	6'0'	176	7/10/1984		1984	
Emmanuel Arceneaux	Wide Receiver	Active	6'2'	210	9/17/1987		1987	
Bernard Berrian	Wide Receiver	Active	6'1'	185	12/27/1980		1980	
Percy Harvin	Wide Receiver	Active	5'11'	192	5/28/1988		1988	
Sidney Rice	Wide Receiver	Active	6'4'	202	9/1/1986		1986	
Visanthe Shiancoe	Tight End	Active	6'4'	250	6/18/1980		1980	
Jim Kleinsasser	Tight End	Active	6'3'	272	1/31/1977		1977	

More Complicated
Some manipulations such as transpose are impossible

One of solutions

To manipulate
To analyze
To visualize

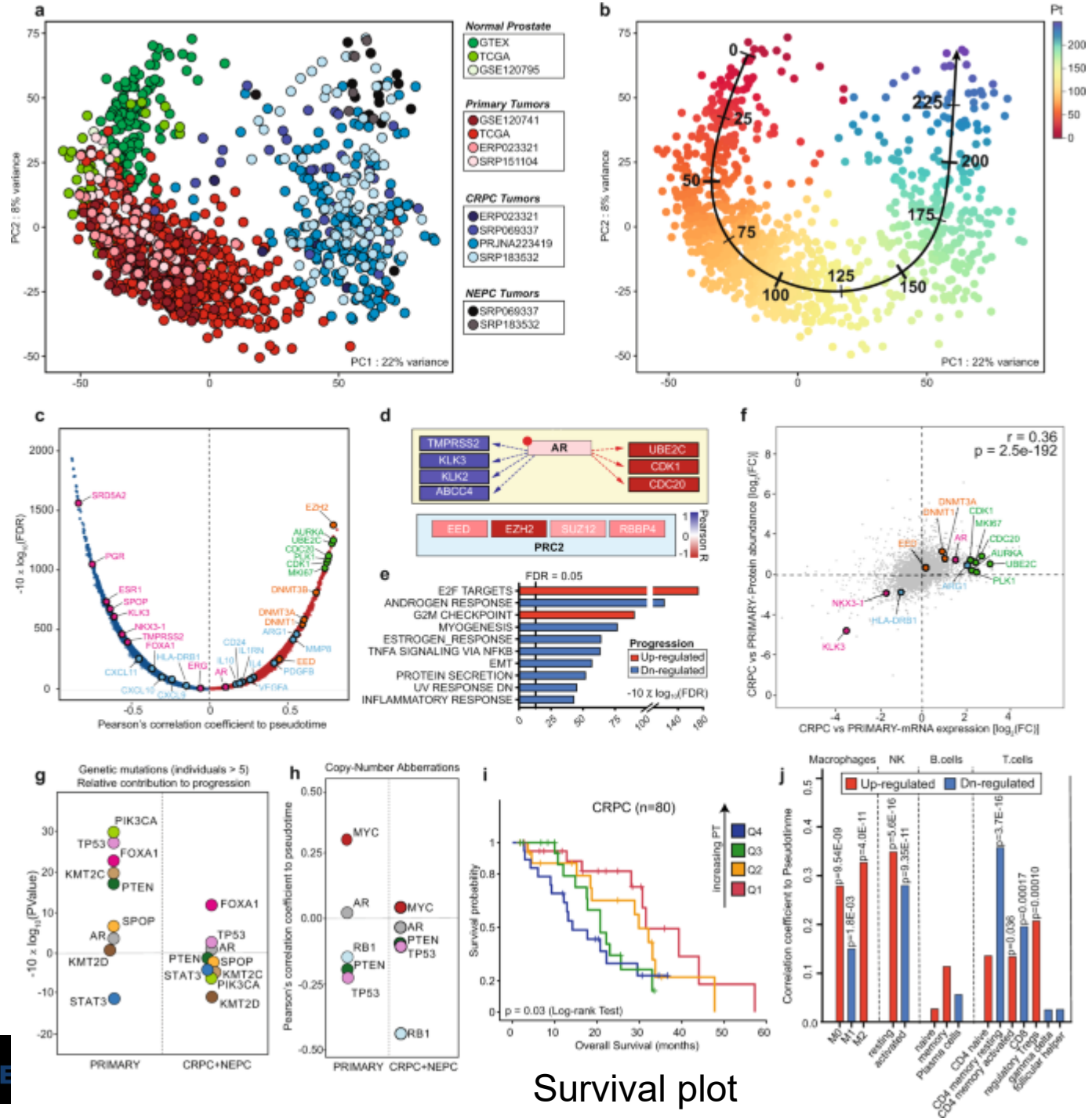


Heatmap

Multilayer-Circular segmented plot

Vasaikar, S.V., Savage, A.K., Gong, Q. *et al.* A comprehensive platform for analyzing longitudinal multi-omics data. *Nat Commun* **14**, 1684 (2023).

Volcano



Survival plot

PCA/tSNE plot

Scatter plot

Bar chart

Bolis, M., Bossi, D., Vallerga, A. *et al.* Dynamic prostate cancer transcriptome analysis delineates the trajectory to disease progression. *Nat Commun* **12**, 7033 (2021). https://doi.org/10.1038/s41467-021-26840-5

What we will learn

- Analyzing omics data quantification

  ..we already taught it in our previous session..

- Advance analytics and graphic visualizations

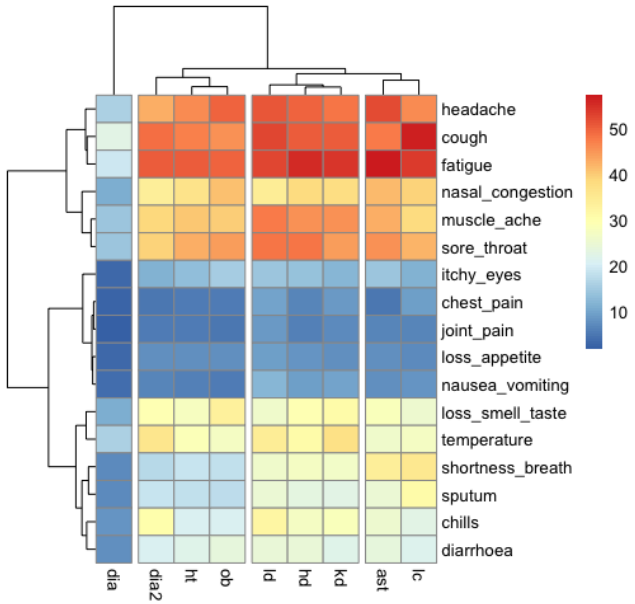
  ..it's up on the character of your data, instead, after this course, you will know tutorials and how to deal with the data by yourself..

- **Basic programming in R**
- Some **basic operations** to look through your data before intensive analysis and visualization
- Some **publication-quality plots** (originated from omics data, in particular transcriptomics, metagenomics, other omics are applicable)

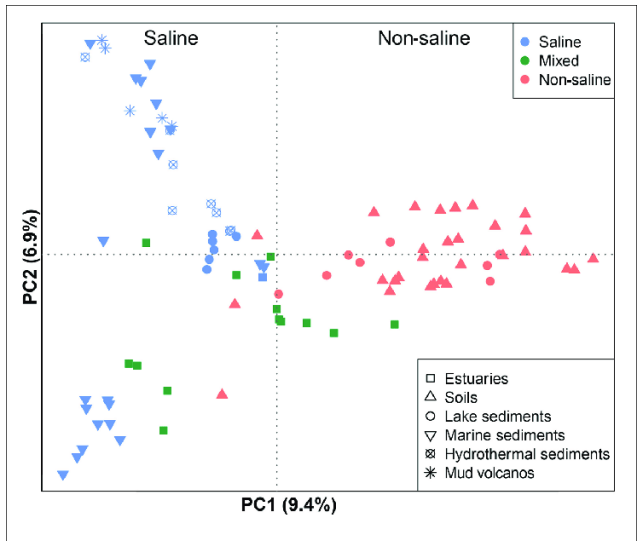
Today

- Basic visualization
- Data manipulation
- Omics data visualization (generals; transcriptomics)
- Omics data visualization (metagenomics)

Omics data visualization (1)

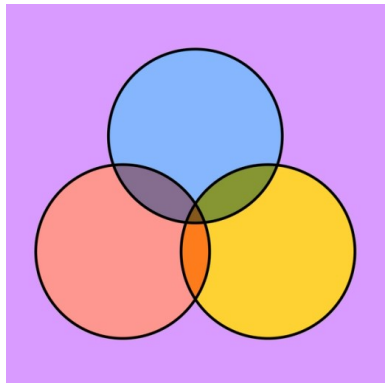


Hierarchical Clustering; **Heatmap**

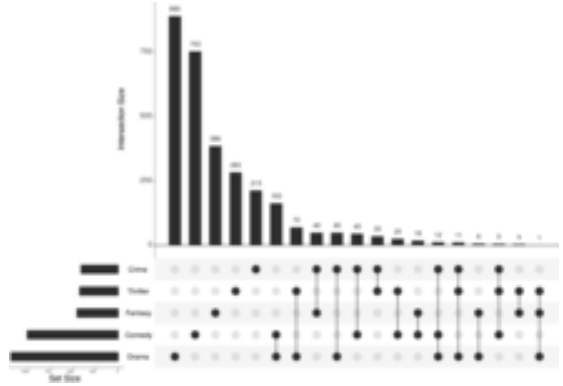


PCA or **PCoA plot**

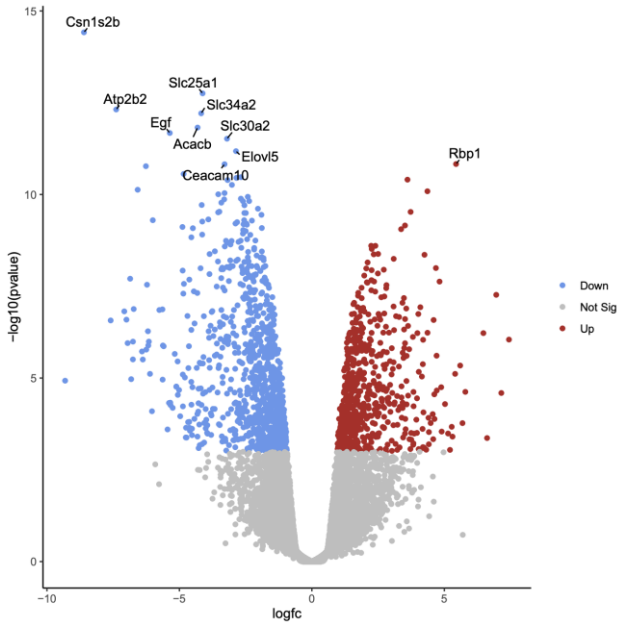
Intersection data



Vann diagram



Upset plot



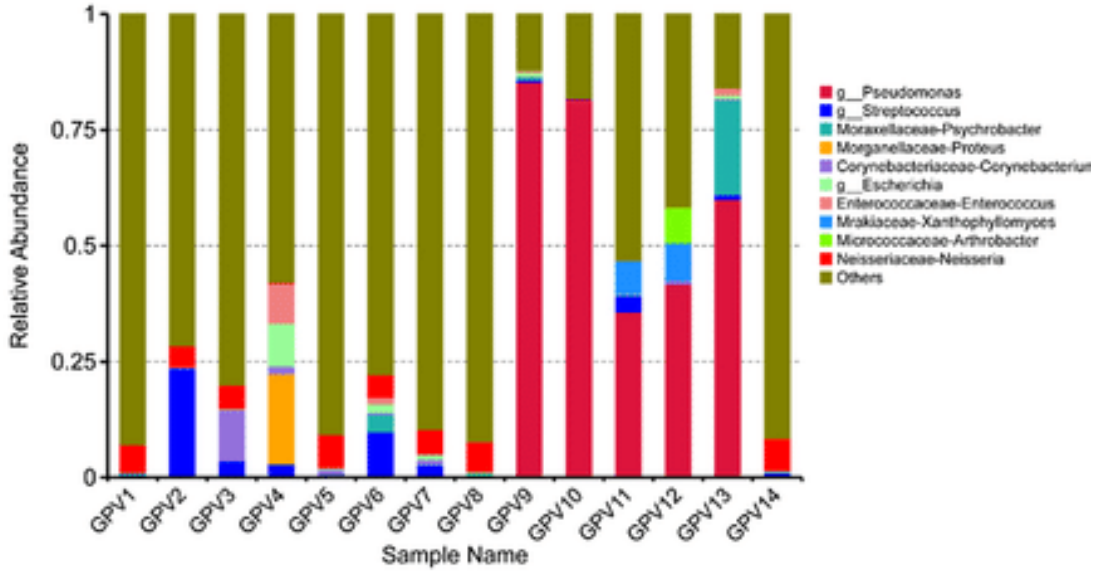
Volcano plot

Omics data visualization (1)

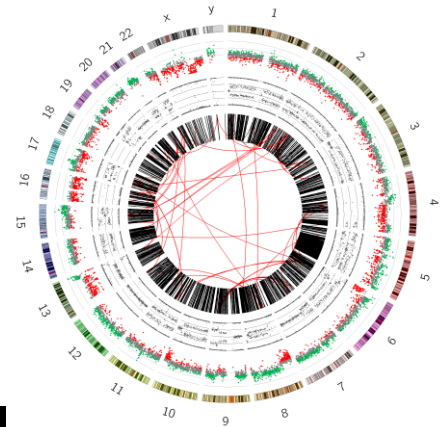
Functional annotation visualization

KEGG, COG, and GO
Bar chart

Omics data visualization (2)

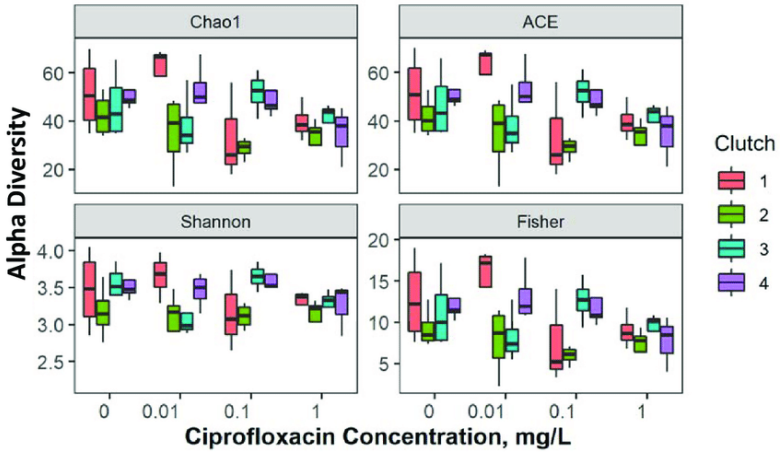


Stacked bar chart; Abundance of Features

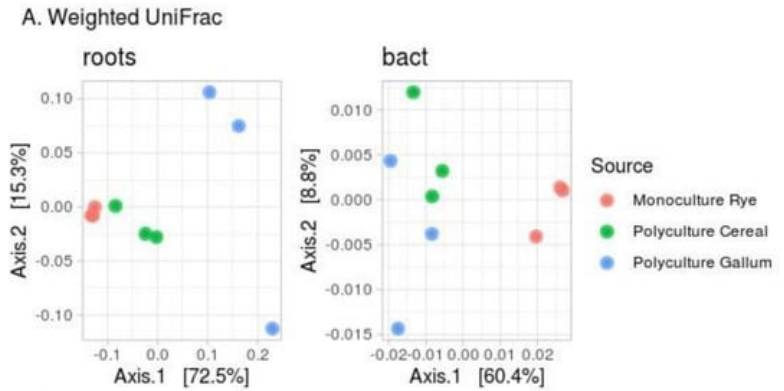


Optional with circus plot (Proksee)

Diversity indices



Alpha diver; **boxplot, dotplot**



Beta diver; **PCA**

	Publisher				
	PLOS	MDPI	BMC	Cell Press	Frontiers
File format	TIFF or EPS	TIFF, JPEG, EPS and PDF	TIFF, JPEG, EPS, PDF, PNG,BMP, PowerPoint	TIFF, JPEG, EPS, and PDF	TIFF, JPEG, EPS
Resolution (dpi)	300 - 600	at least 600	at least 300	300 - 1000	300
Dimensions maximum (W*H)	2250*2625	-	1200px (W)	-	-
Dimensions minimum (W*H)	789 (W)	1000*1000	600px (W)	-	-
File size	<10 MB	120 MB (in total)	10 MB	20 MB, 3 MB (PDF)	-
link	Publisher's site	Publisher's site	Publisher's site	Publisher's site	Publisher's site

Remarks: Depending on the journal of choice, graphic requirements may vary. It may be necessary for authors to check the journal's website before preparing graphics.

shutterstock.com · 1263463690

FEEDBACK

https://certificate.sci.psu.ac.th/site/a#/publish/course/41/question

Molecular Biotechnology and Bioinformatics

สถานวิจัยจีโนมและชีวสารสนเทศ มหาวิทยาลัยสงขลานครินทร์

Agricultural Biotech & Bioinformatics Lab

Bioinformatics Workshop EP3.

GWAS and marker selection in agriculture

3-5 July 2023

PSU

AG-BIO

CENTER FOR AGRICULTURAL BIOTECHNOLOGY

KU KAMPANGSAN CAMPUS

FACULTY OF NATURAL RESOURCES

PRINCE OF SONGKLA UNIVERSITY

Agricultural Biotech & Bioinformatics Lab

SCIENCE

Agricultural Biotech & Bioinformatics Lab

ABBLab Sequencing Platform

Metagenomic/Amplicon sequencing
Microorganism genome sequencing
Genotyping by sequencing

Oxford Nanopore MinION

[Agricultural Biotech & Bioinformatics Lab, PSU](#)

PSU SCIENCE

Ponsit.sat@gmail.com , ponsit.s@psu.ac.th