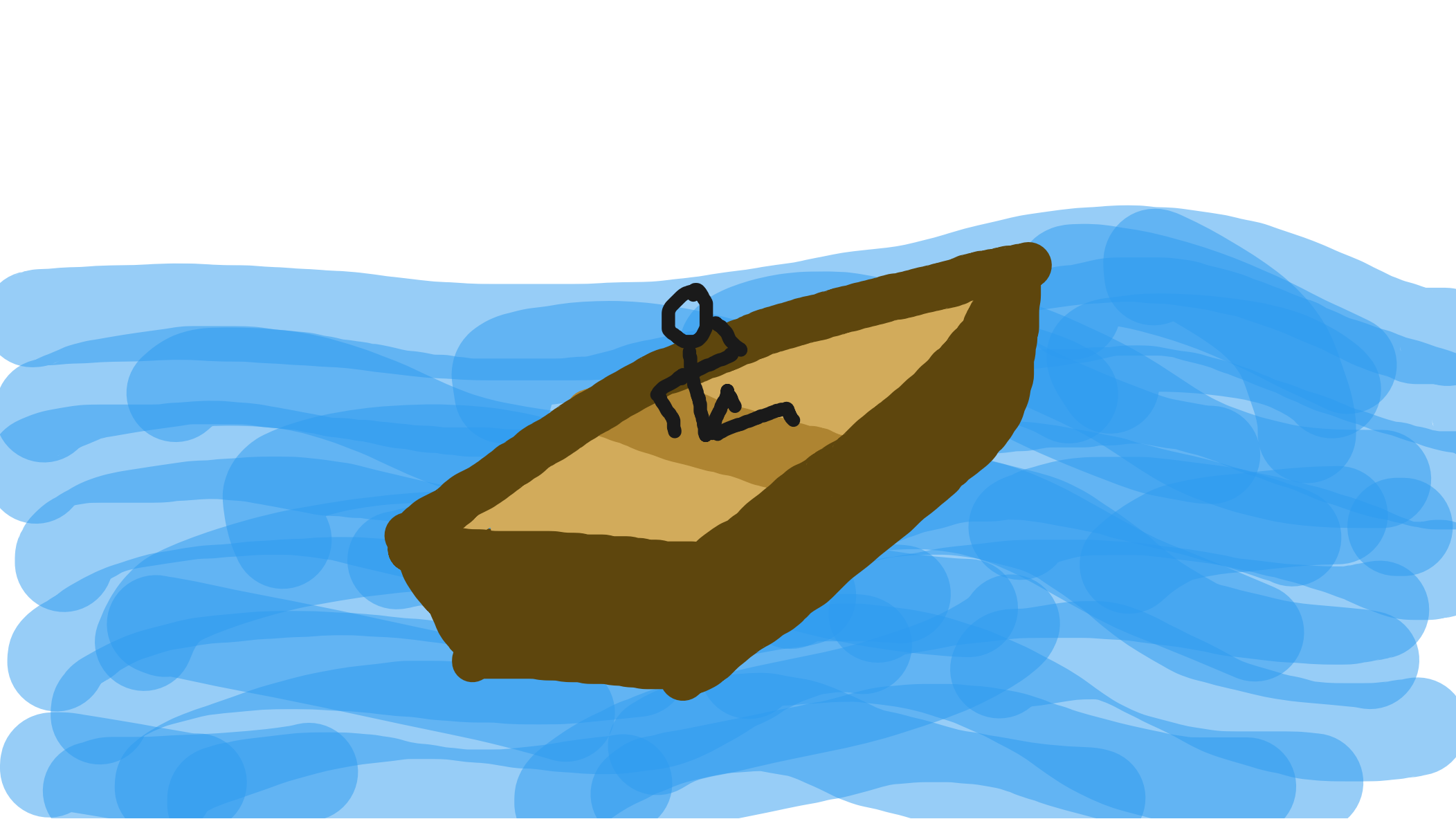
Data data everywhere,
No time to think

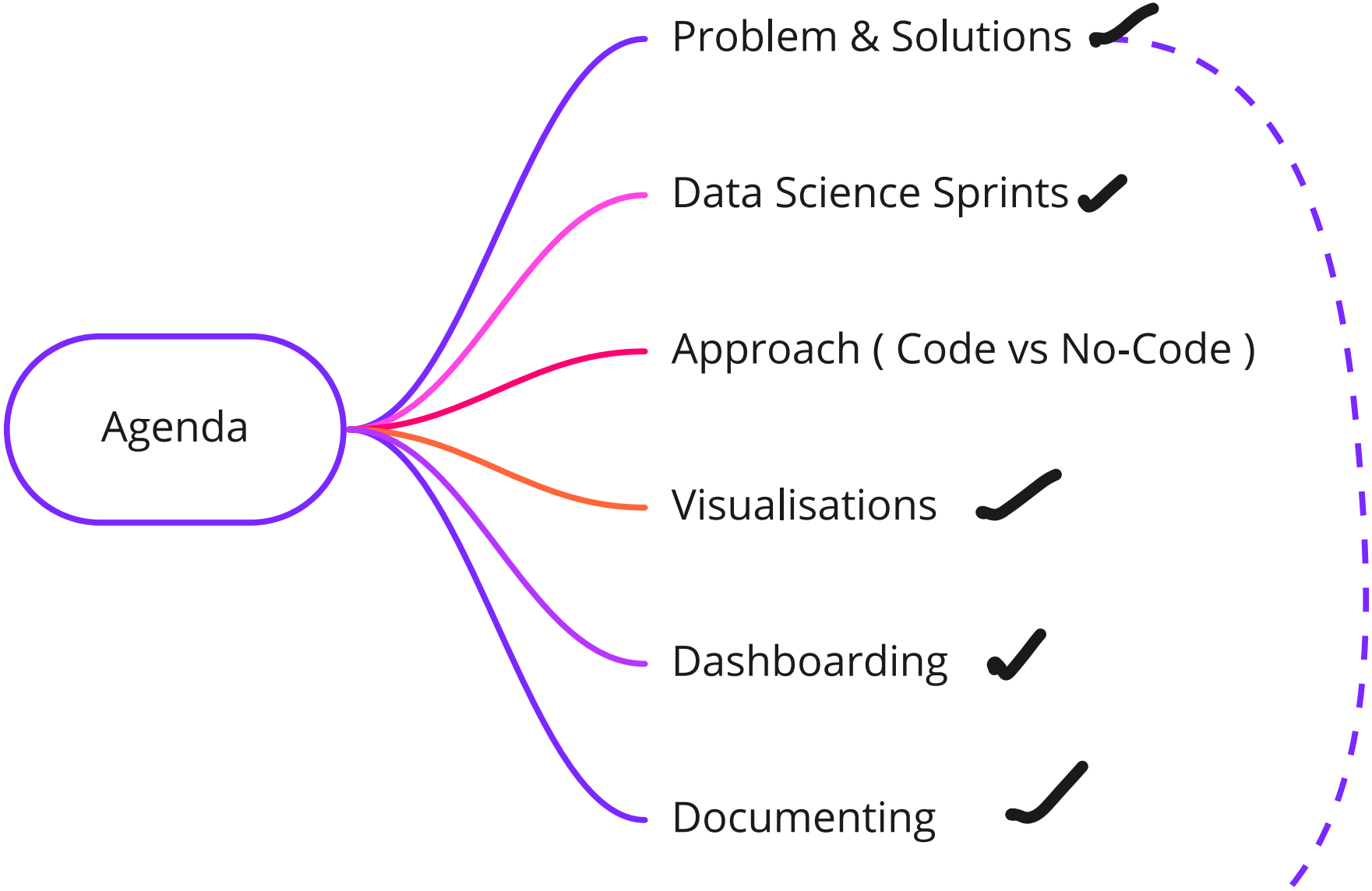
Aman Sharma

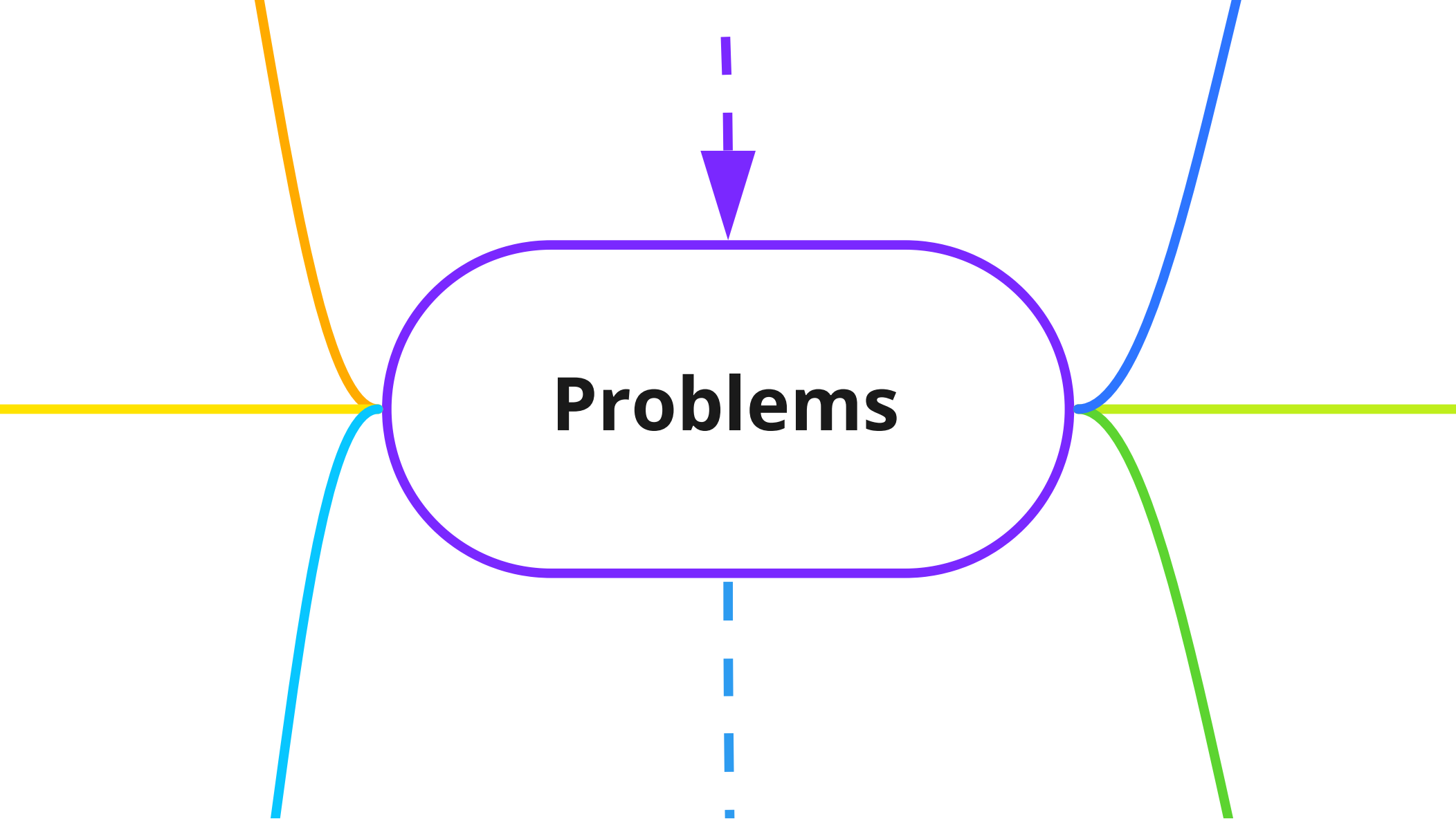
AKA @amanintech

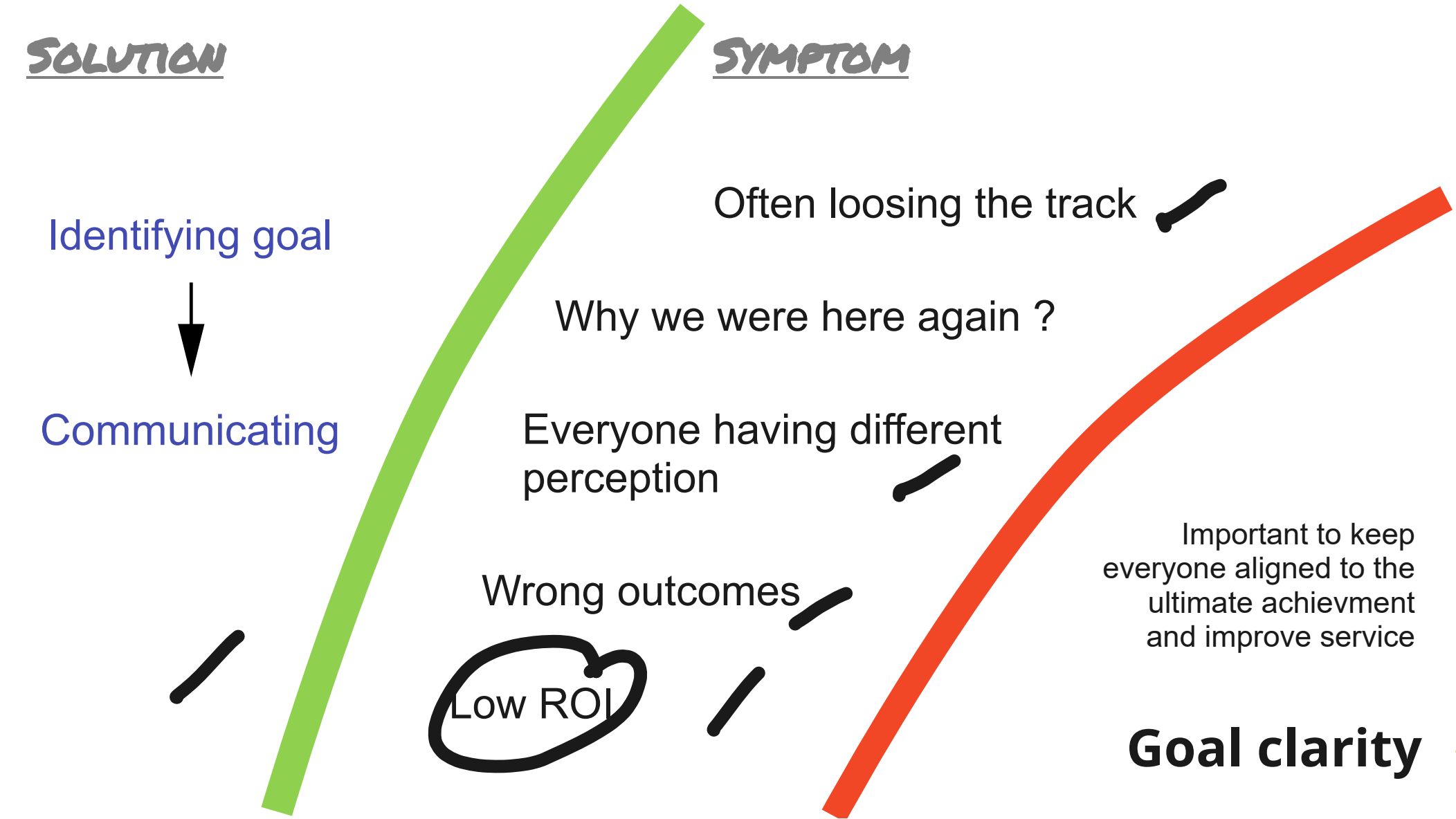
- Co-founder & CTO **twimbit**
- Member **deeplearning.ai**
- Mentor | Entreprenuer | Tech Generalist


Whats
difference
between Data
and Insights ?




Agenda
Problem & Solutions
Data Science Sprints
Approach (Code vs No-Code)
Visualisations
Dashboarding
Documenting


Problems


Solution
Symptom
Identifying goal
Communicating
Often loosing the track
Why we were here again ?
Everyone having different perception
Wrong outcomes
Low ROI
Important to keep everyone aligned to the ultimate achievment and improve service
Goal clarity

Solution

Better project structure

↓

Less Deliverables

↓

Strict timelines

Symptom

Missing deadlines

Low yield results

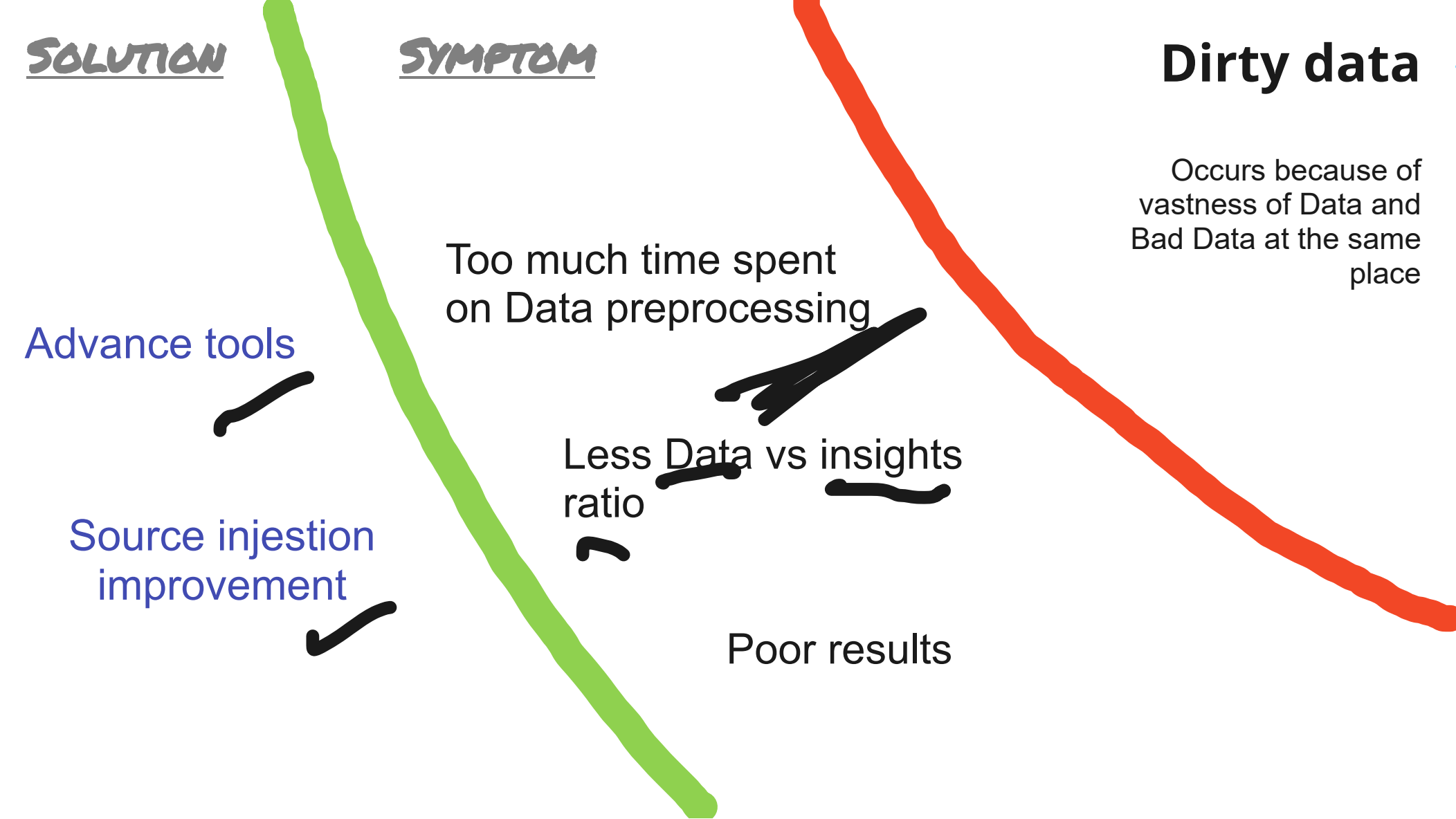
Less resources ?

Less resources ?

Reason for chaotic situation and abandonment

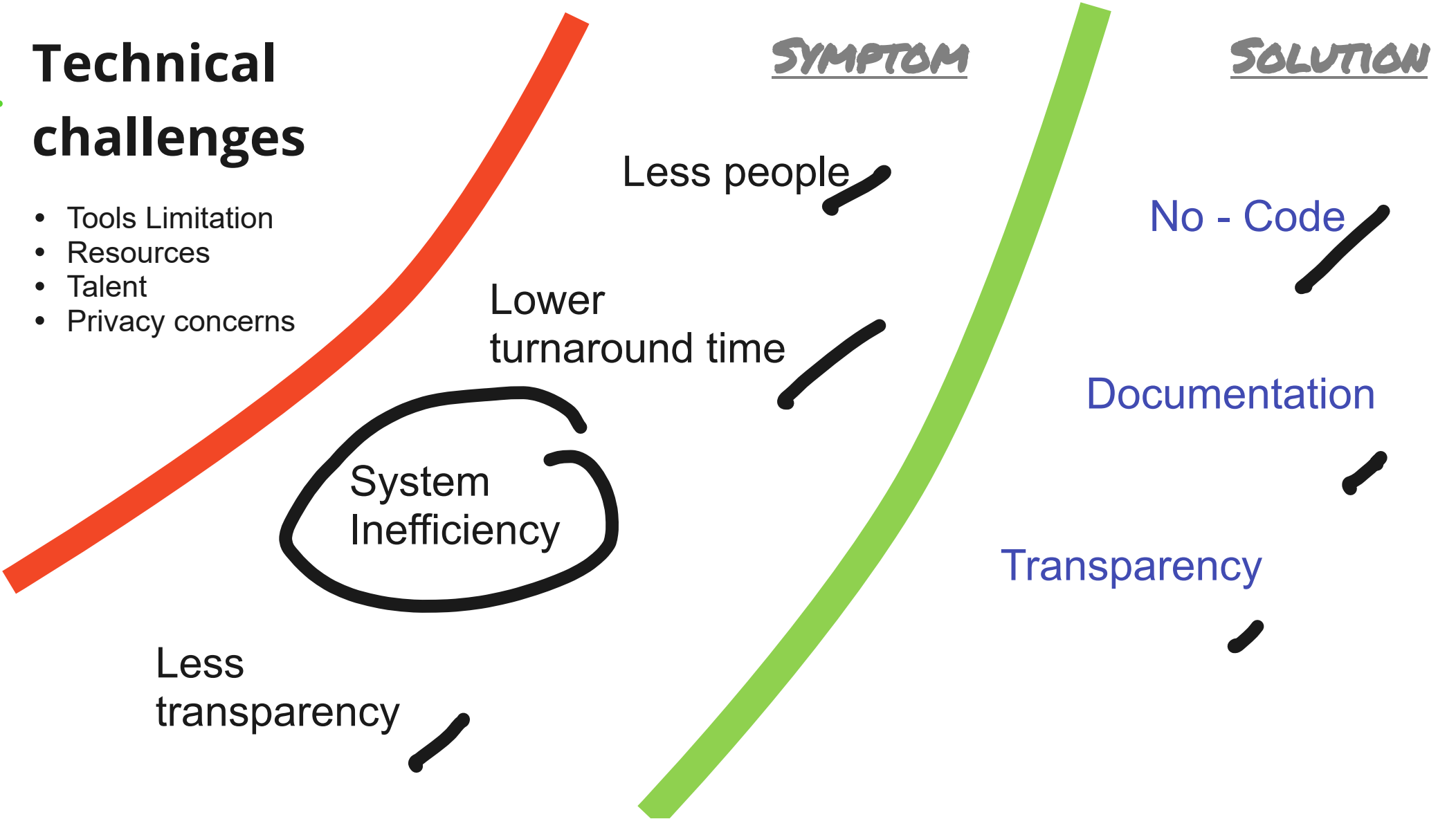
Poor planning

Leads to wastage of time and resources


Solution
Symptom
Dirty data
Occurs because of vastness of Data and Bad Data at the same place
Too much time spent on Data preprocessing
Advance tools
Less Data vs insights ratio
Source injestion improvement
Poor results

Technical challenges

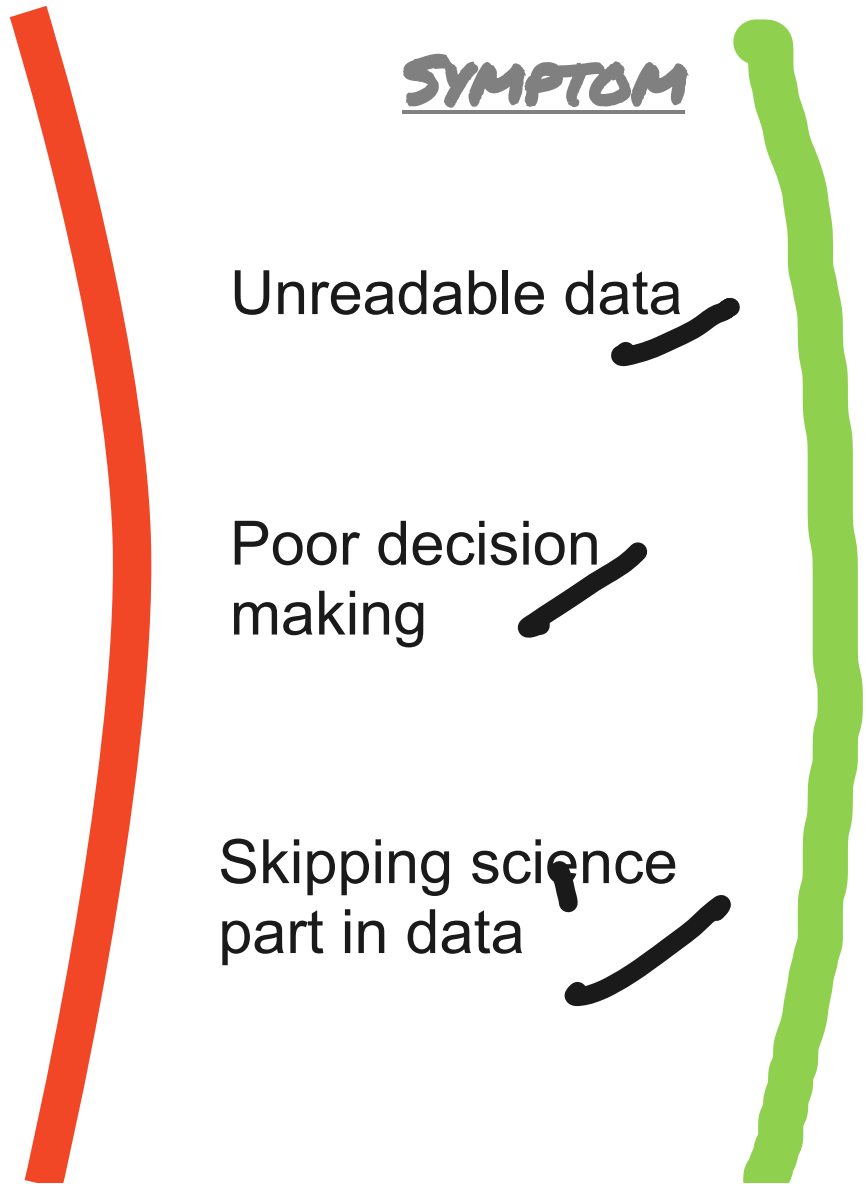
- Tools Limitation
- Resources
- Talent
- Privacy concerns



Data Science limitedness to only technical people

Complexity

Insufficiency in Data Representation due to information being too un-understandable



SYMPTOM

Unreadable data

Poor decision making

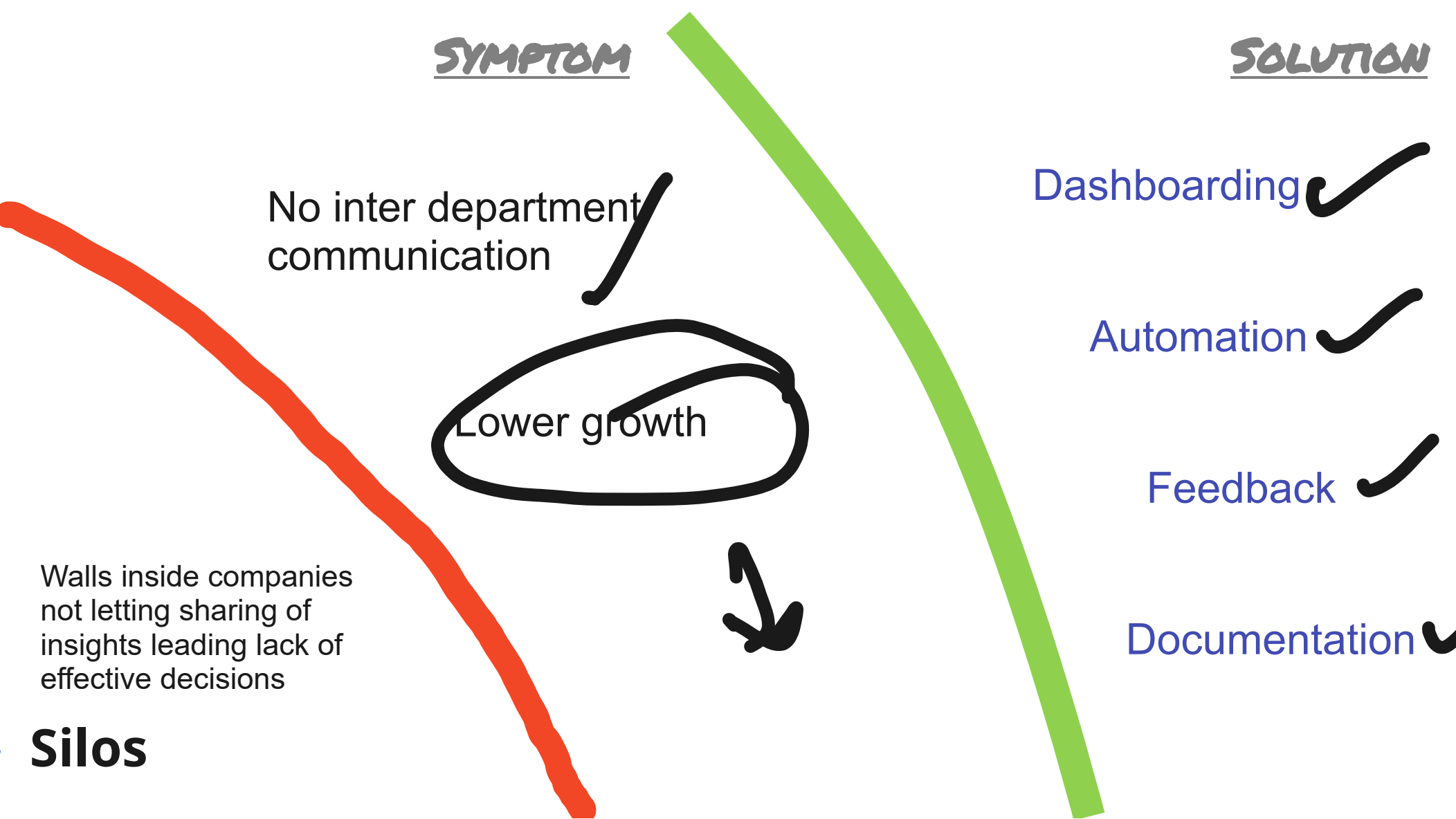
Skipping science part in data

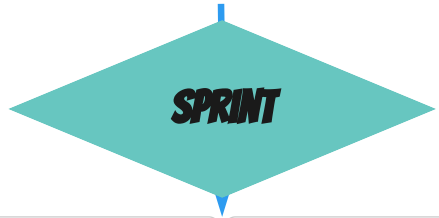
SOLUTION

No-Code

Better visualisation

Feedback


Symptom
Solution
No inter department communication
Lower growth
Walls inside companies not letting sharing of insights leading lack of effective decisions
Silos
Dashboarding
Automation
Feedback
Documentation



Kanban

Clear Goal | 4

- Introduce the project
- Set a long term Goal
- Set Sprint Questions
- Make a Map

Plan Well | 5

- Talk With Experts
- Pick a Target
- How might we notes
- Pick a Approach
- Create flow diagram

Execute | 4

- Set Deliverables
- Set Timeline
- Divide Tasks
- Meet Regularly

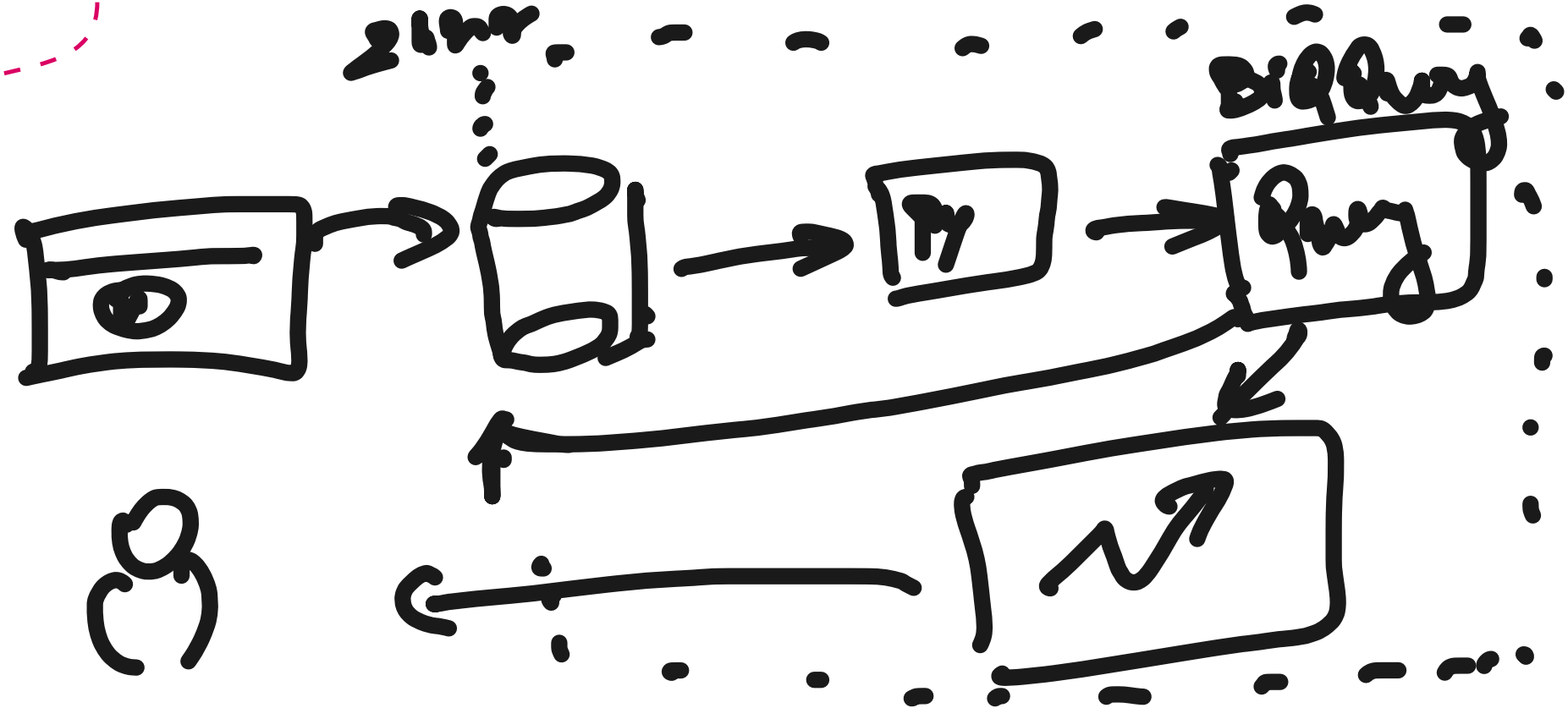
Test and improvise | 5

- Create an MVP / Test Report
- Present findings
- Collect feedbacks
- Improvise
- Document Learnings

For more info - SPRINT book by Jake Knapp

DIAGRAMMING

- Trying to get everyone thoughts clear
- Setting realistic acheivables + timelines
- Estimating resources



APPROACH

CODE

Pros

- Flexibility
- Scalability
- High function availability

Cons

- Technical proficiency
- Talent
- Silos
- Model Complexity

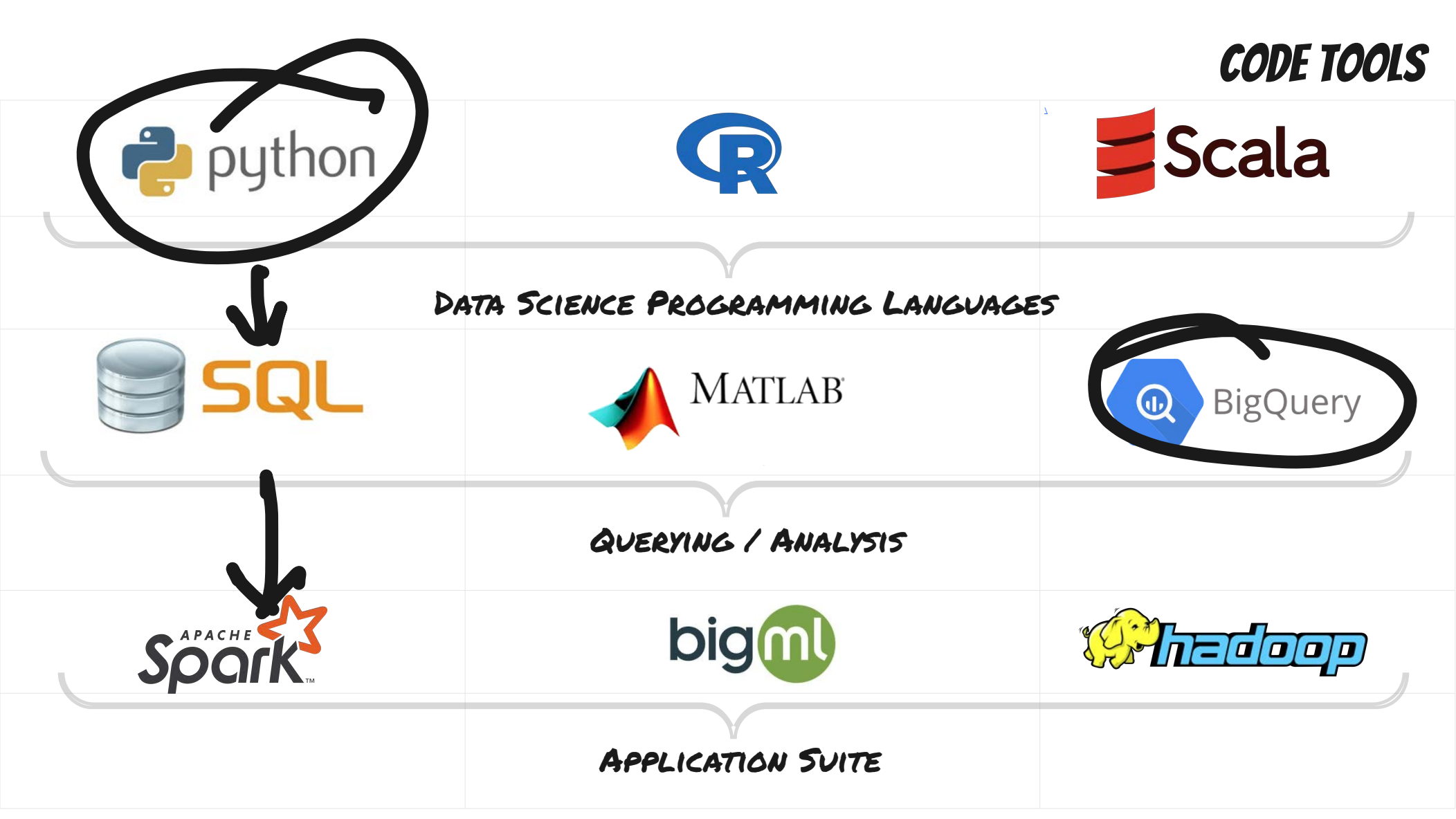
NO-CODE/ LOW CODE

Pros

- Speed
- Low learning curve
- Fun and Engaging
- Increase productivity
- Openness

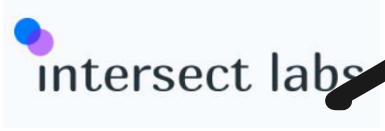
Cons

- Flexibility
- Source Choices
- Less options
- Dependency
- Scalability


CODE TOOLS
python
R
Scala
Data Science Programming Languages
SQL
MATLAB
BigQuery
Querying / Analysis
APACHE Spark
bigml
hadoop
Application Suite

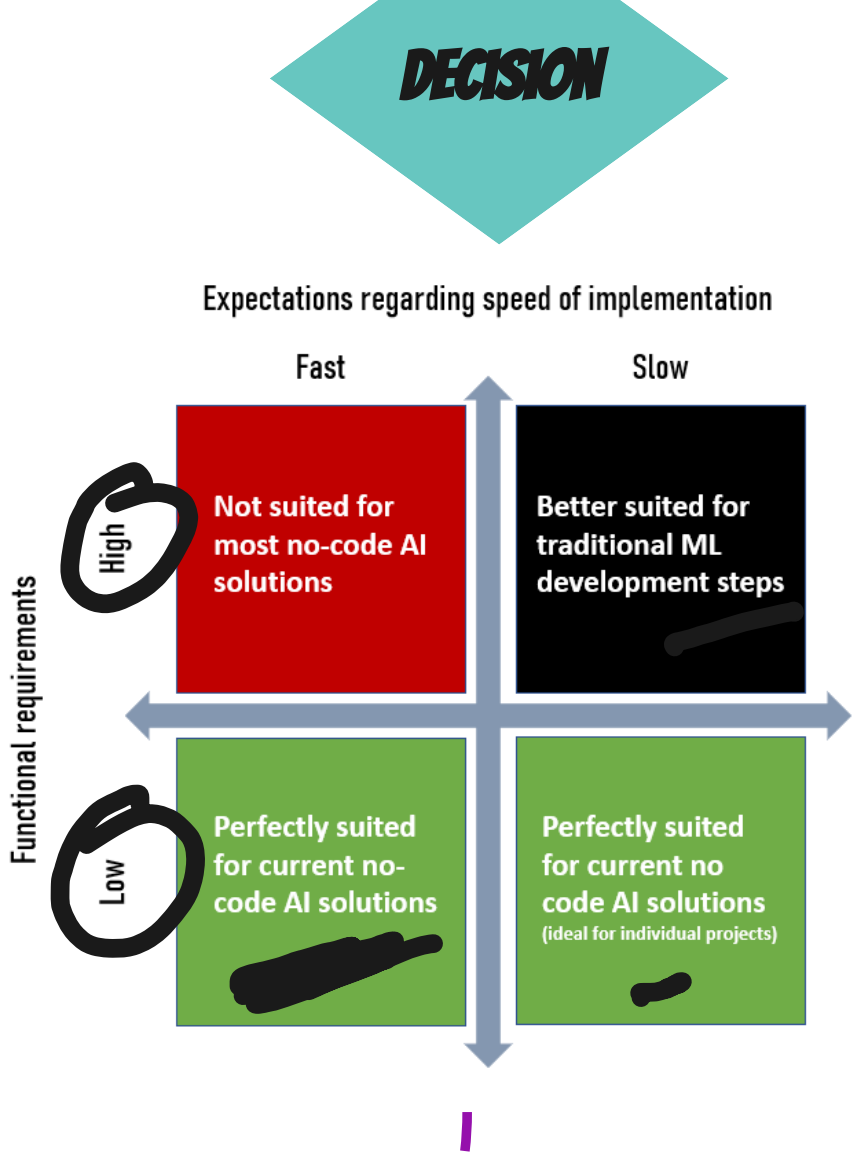
NO- CODE/ LOW CODE TOOLS

Easy to create dashboards and reports



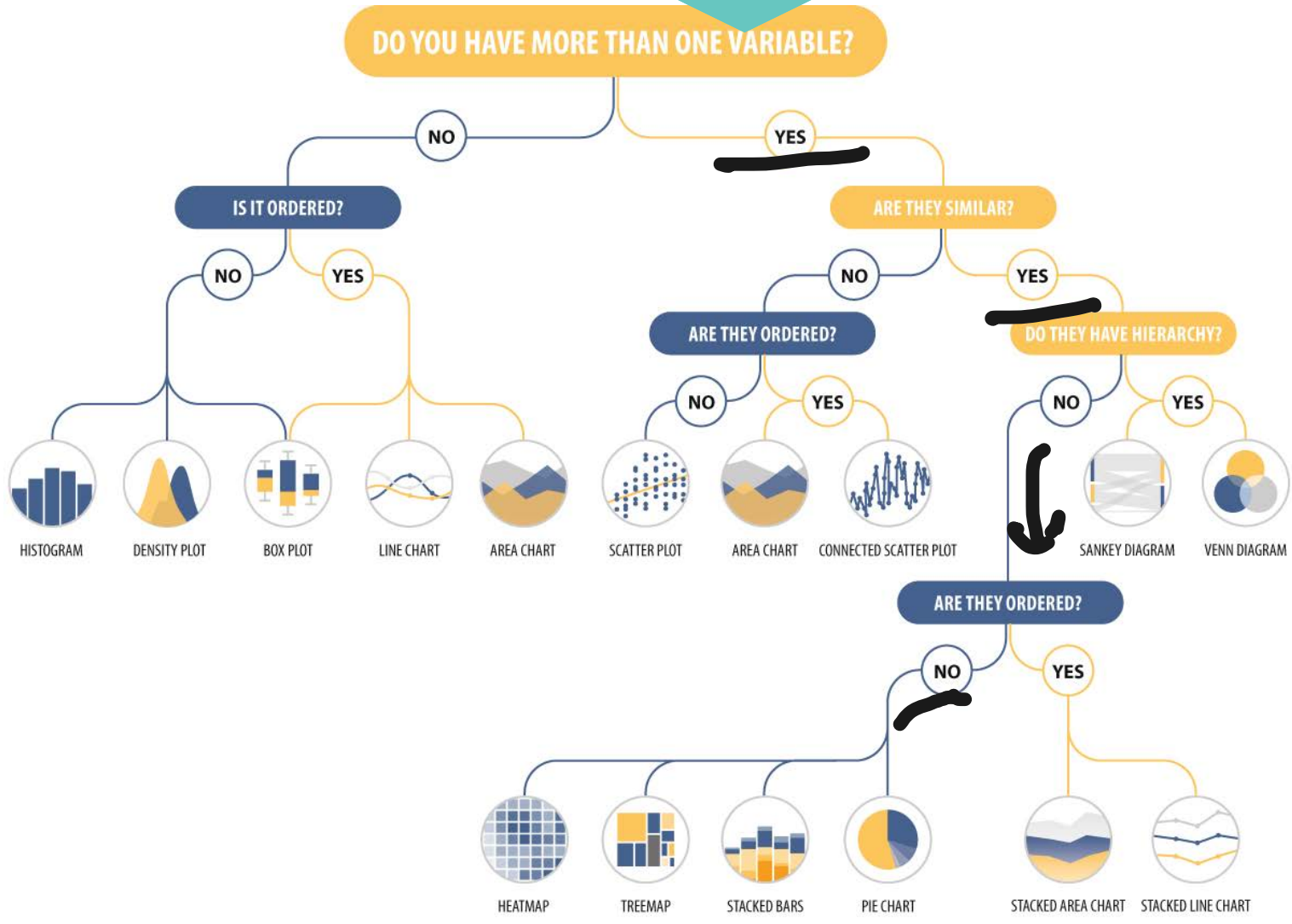
Build and Automate data Science flows

End to end Data science flow


DECISION
Expectations regarding speed of implementation
Fast
Slow
Functional requirements
High
Low
Not suited for most no-code AI solutions
Better suited for traditional ML development steps
Perfectly suited for current no-code AI solutions
Perfectly suited for current no code AI solutions
(ideal for individual projects)

VISUALISATION

more about charts



D3

plotly

matplotlib

Nuts and Bolts of Chart Types

Created by
online-behavior.com

HISTOGRAM

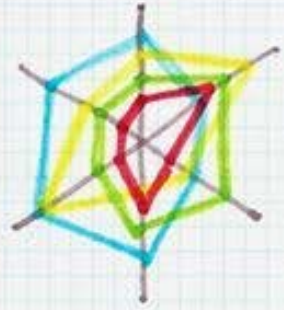
Chart used by responsible analysts who understand the power of segmentation and the sadness that comes from aggregating data.

BAR CHART

Safe choice. But make sure you read Stephen Few before you show the chart to your boss, it will increase the probability of getting a raise.

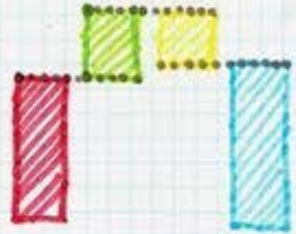
BUBBLE CHART

If you manage to extract insights from this graph your name is Hans Rosling.



RADAR CHART

If you want to build a complex model around which you have done a ton of research, that's your choice (but only PhDs will understand you).



WATERFALL CHART

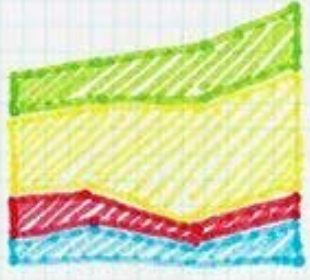
Perfect if you want to hide information or misguide other people. Seriously, can you trust a chart that is also known as a "Flying Bricks Chart" or "Mario Chart"? No.

PIE CHART

Extremely useful when creating a well designed document that is intended to people that will not read the data (e.g. management)

LINE CHART

Useful to show trends, especially upwards (for downwards trends people tend to use more obscure charts, like the waterfall chart, see below)



AREA CHART

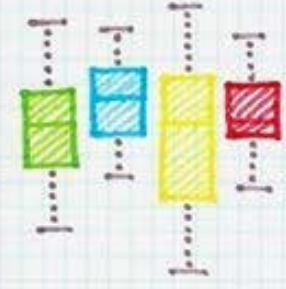
Please don't use this chart, I beg you! And please buy one of Tufte books.

TREE MAP

I have seen trees and I have seen maps, but how exactly this map is a combination of both? If you use it, good for you.

SCATTERPLOT

Very useful to find outliers, just like the people that commonly create them: human beings that finished their PhD in math by the age of 16.



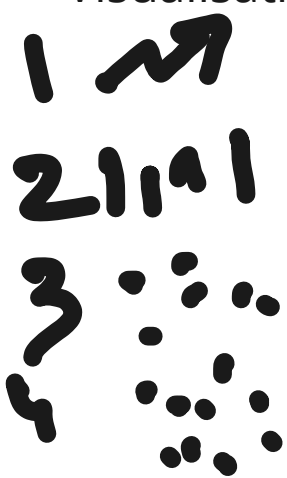
BOX PLOT

This one is for pros. If you use it successfully, you will get a seat in heaven between Ronald Fisher and Johann Carl Friedrich Gauss.

DASHBOARDING

Effectively Sharing findings and work with everyone

Collect Visualisation → organise → Set a scheduled Automation → Share access

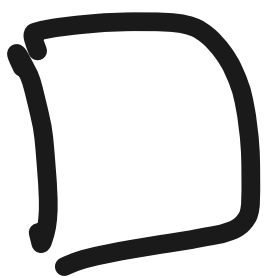


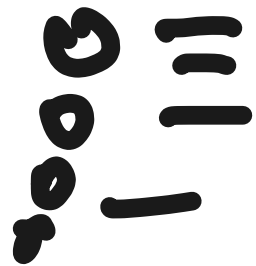
Apache Superset™ Apache Airflow gramex plotly | Dash amazon QuickSight

DOCUMENTING

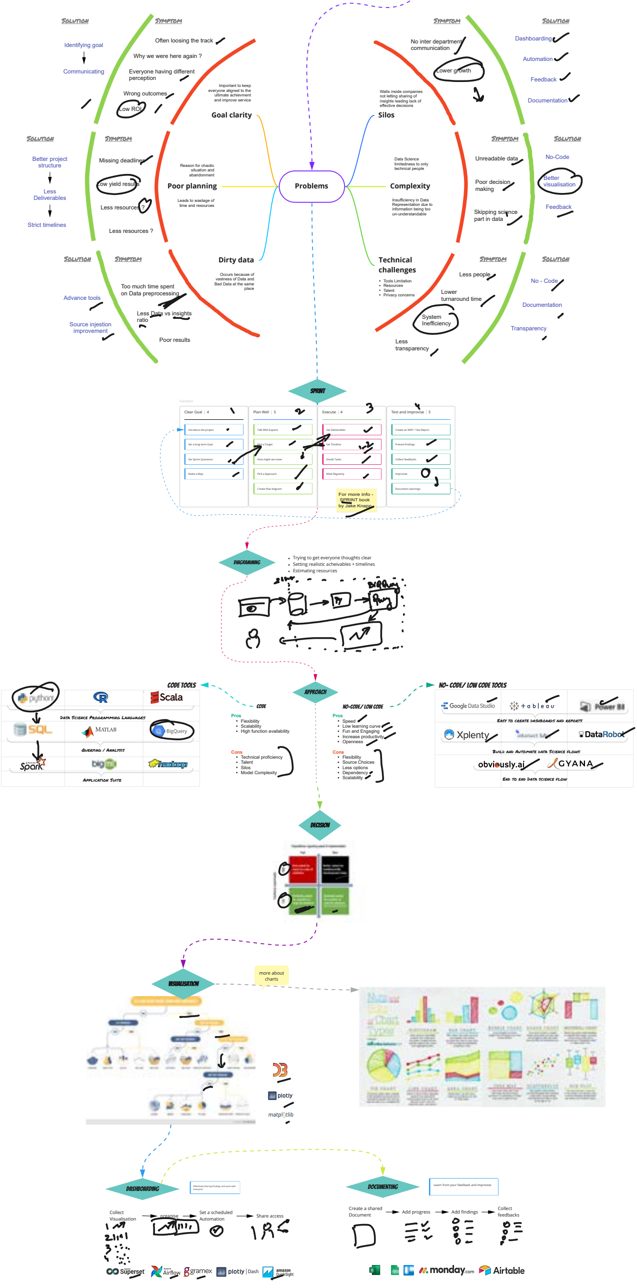
Learn from your feedback and improvise

Create a shared Document → Add progress → Add findings → Collect feedbacks



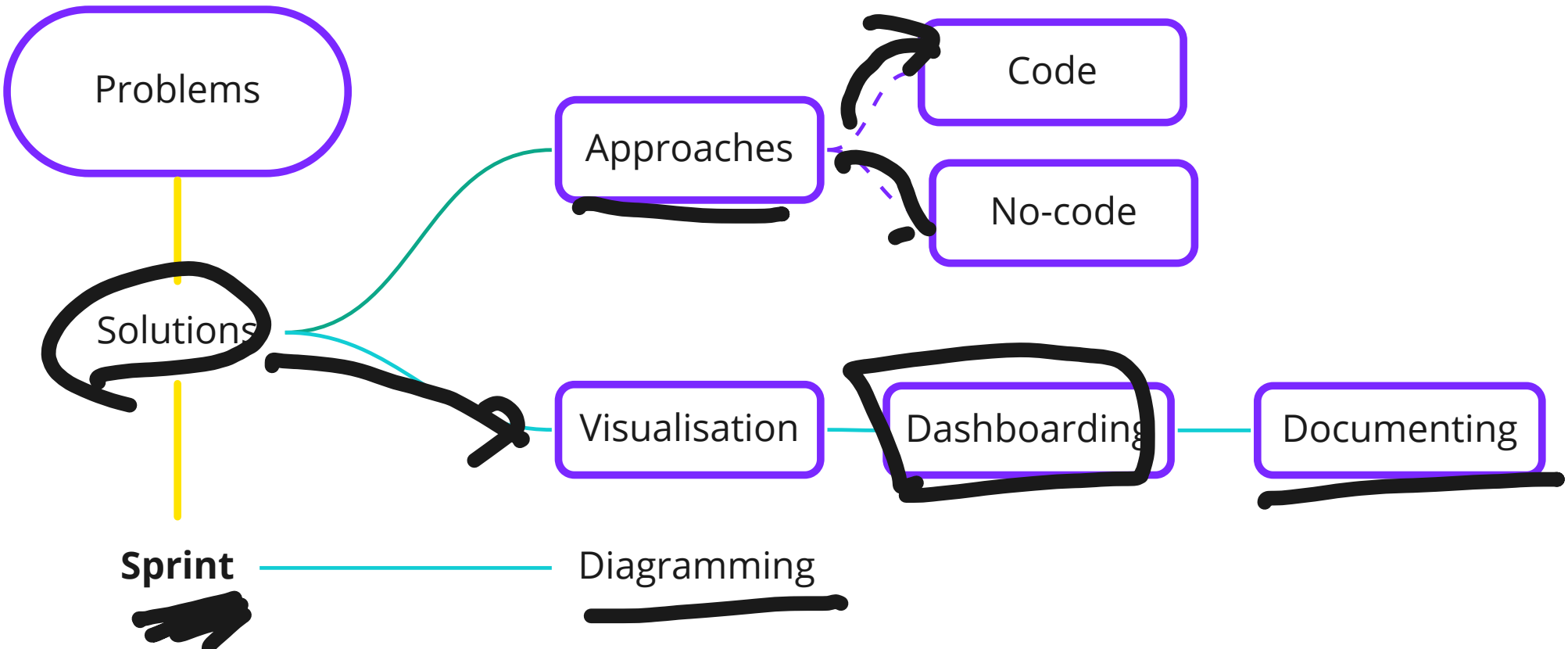


monday.com Airtable



Problems
Goal clarity
Important to keep everyone aligned to the ultimate achievment and improve service
Symptom
Solution
Often loosing the track
Why we were here again ?
Everyone having different perception
Wrong outcomes
Low ROI
Identifying goal
Communicating
Poor planning
Reason for chaotic situation and abandonment
Leads to wastage of time and resources
Missing deadlines
Low yield results
Less resources ?
Less resources ?
Better project structure
Less Deliverables
Strict timelines
Dirty data
Occurs because of vastness of Data and Bad Data at the same place
Too much time spent on Data preprocessing
Less Data vs insights ratio
Poor results
Advance tools
Source injestion improvement
Silos
Walls inside companies not letting sharing of insights leading lack of effective decisions
No inter department communication
Lower growth
Dashboarding
Automation
Feedback
Documentation
Complexity
Data Science limitedness to only technical people
Insufficiency in Data Representation due to Information being too un-understandable
Unreadable data
Poor decision making
Skipping science part in data
No-Code
Better visualisation
Feedback
Technical challenges
Tools Limitation
Resources
Talent
Privacy concerns
Less people
Lower turnaround time
System Inefficiency
Less transparency
No - Code
Documentation
Transparency
Sprint
Kanban
Clear Goal | 4
Introduce the project
Set a long term Goal
Set Sprint Questions
Make a Map
Plan Well | 5
Talk With Experts
Pick a Target
How might we notes
Pick a Approach
Create flow diagram
Execute | 4
Set Deliverables
Set Timeline
Divide Tasks
Meet Regularly
Test and improvise | 5
Create an MVP / Test Report
Present findings
Collect feedbacks
Improvise
Document Learnings
For more info - SPRINT book by Jake Knapp
Diagramming
Trying to get everyone thoughts clear
Setting realistic acheivables + timelines
Estimating resources
BigQuery
Approach
Code tools
Scala
Data Science Programming Languages
SQL
MATLAB
BigQuery
Querying / Analysis
Spark
Application Suite
Code
Pros
Flexibility
Scalability
High function availability
Cons
Technical proficiency
Talent
Silos
Model Complexity
No-code/ Low code
Pros
Speed
Low learning curve
Fun and Engaging
Increase productivity
Openness
Cons
Flexibility
Source Choices
Less options
Dependency
Scalability
No- code/ Low code tools
Google Data Studio
tableau
Power BI
Easy to create dashboards and reports
Xplenty
DataRobot
Build and automate data science flows
obviously.ai
GYANA
End to end data science flow
Decision
Visualisation
more about charts
D3
plotly
matplotlib
Dashboarding
Effectively Sharing findings and work with everyone
Collect Visualisation
organise
Set a scheduled Automation
Share access
Superset
Airflow
gramex
plotly | Dash
amazon QuickSight
Documenting
Learn from your feedback and improvise
Create a shared Document
Add progress
Add findings
Collect feedbacks
monday.com
Airtable

SUMMARY


Problems
Solutions
Sprint
Approaches
Code
No-code
Visualisation
Dashboarding
Documenting
Diagramming

Data data everywhere,
No time to think

→ By Aman Sharma

AKA @amanintech

THANKS FOR YOUR TIME !!