[G0W36A] Capita Selecta in Statistics
Visual Data Analysis

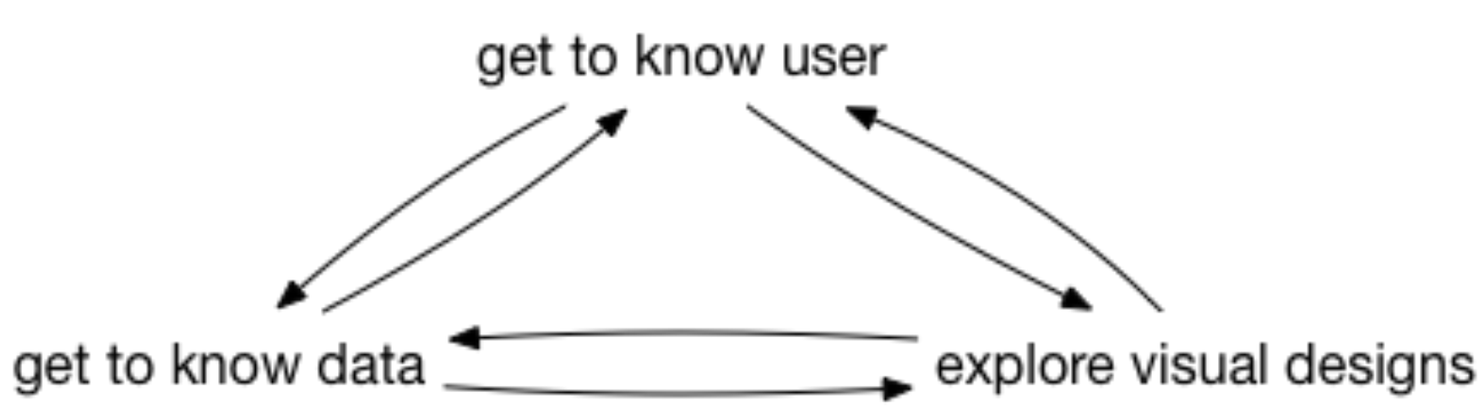
Prof Jan Aerts
Visual Data Analysis lab, ESAT/STADIUS
Faculty of Engineering
KU Leuven

@jandot - jan.aerts@esat.kuleuven.be - http://vda-lab.be

TAs: Daniel Alcaide, Jansi Thiyagarajan, Houda Lamqaddam
(firstname.lastname@esat.kuleuven.be)

(several parts as described by Francis Rowland, EBI)

The design process



The user

Find the why

- what they *want* != what the *need* => need to find **underlying goals**
- e.g. let them **imagine** what they could do if some technologies were available that are (still) science-fiction (e.g. nanobots in blood; Gaviscon commercial http://m.youtube.com/watch?v=_skKmcLdyVQ) => helps to identify underlying **assumptions**
- additional methods, e.g. **card sorting**
- if possible: tape the discussion (w/ agreement)
- ask “why?” 3 times

Nanobot Specification Sheet (version 2253)

Requested by: Jan

Nanobot name: JAN-001

Nanobot type: ☒ sensor (measure/monitor/detect)
☐ execute

Type of information to be gathered (general description; 1 sentence):

virus entered cell

Sensor type: ☐ continuous; polling frequency: _____
☒ triggered; what trigger? virus entered
☐ once

To be administered: ☒ at birth
☐ ad hoc; when: _____

Sketch of dummy data format:

- 3D position cell
- tissue
- which virus

Nanobot Specification Sheet (version 2253)

Requested by: Jan

Nanobot name: JAN-002

Nanobot type: ☒ sensor (measure/monitor/detect)
☐ execute

Type of information to be gathered (general description; 1 sentence):

expression level of p53 (actual copies of protein)

Sensor type: ☒ continuous; polling frequency: 5 min
☐ triggered; what trigger? _____
☐ once

To be administered: ☒ at birth
☐ ad hoc; when: _____

Sketch of dummy data format:

timestamp	tissue	expression level
22531120932	liver	5,372,923
22531120932	heart	2,793,529
⋮		

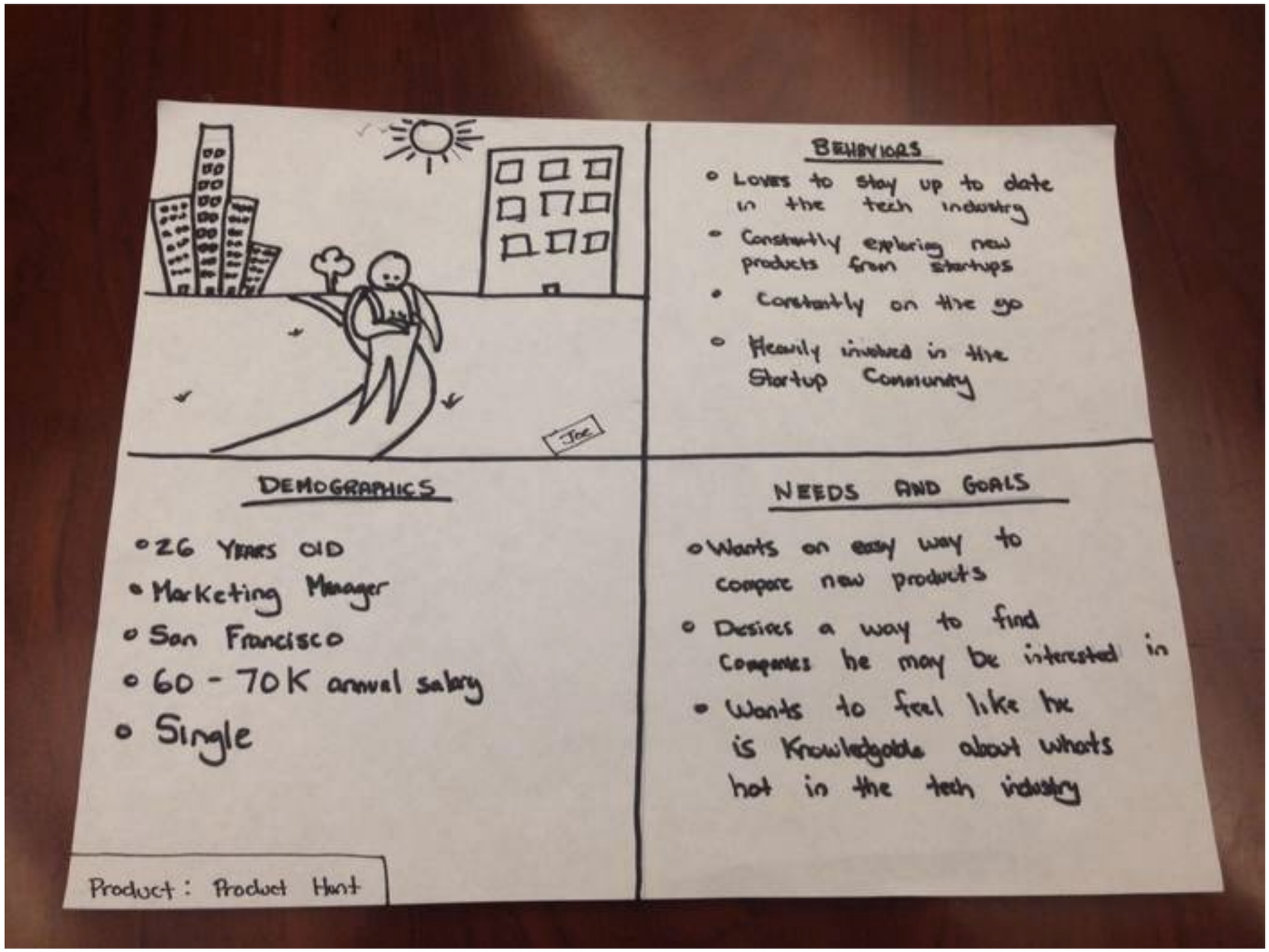
Statement of goals

- Based on this discussion, state the specific goals of the user (= task abstraction as described by T. Munzner), e.g.
    - “Show the relationship between A, B and C across X and Y from m to n”.
    - “Identify instances of A that have a value larger than x”.
    - …

Proto-persona

- Based on this task analysis, create one or (probably) more porto-personae, describing their name, behaviours and characteristics, needs or pain points, and what would help them. Who would be the actual users, and why do the need it?

Yun, clinical researcher  www.edmonton.ca	**Behaviour and characteristics** Yun spends about half of her time in the lab but uses bioinformatics tools and software to analyse and process disease-related data. Often stays late (experiments!)
Needs & pain points Often doesn't have time to remember how to use certain software Feels overwhelmed by latest huge datasets Finds some visualisations overly-complex	**Would be served by…** Summary reports of data with highlights Focus on certain genes by default Add and mix data, perhaps in layers



http://www.anthonycreyes.com/persona-to-prototype-product-hunt/

[Activity] Proto-persona & problem statements (in group)

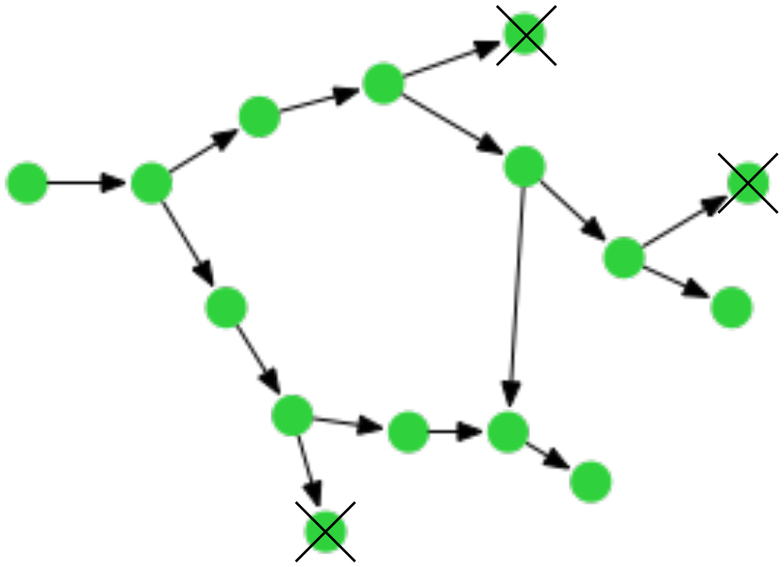
- dataset = flight data

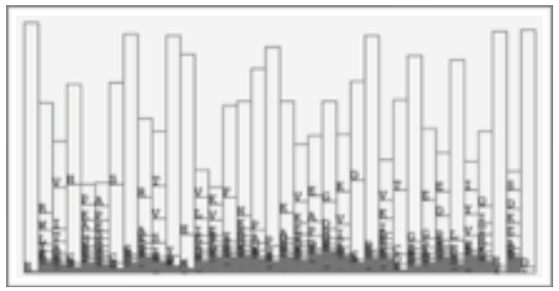
```
from_airport,from_city,from_country,from_long,from_lat,to_airport,to_city,to_country,to_long,to_lat,airline,airline_country,dist
ance
Balandino,Chelyabinsk,Russia,61.838,55.509,Domododevo,Moscow,Russia,38.51,55.681,Aerocondor,Portugal,1458
Balandino,Chelyabinsk,Russia,61.838,55.509,Kazan,Kazan,Russia,49.464,56.01,Aerocondor,Portugal,775
Balandino,Chelyabinsk,Russia,61.838,55.509,Tolmachevo,Novosibirsk,Russia,83.084,55.021,Aerocondor,Portugal,1341
...
```

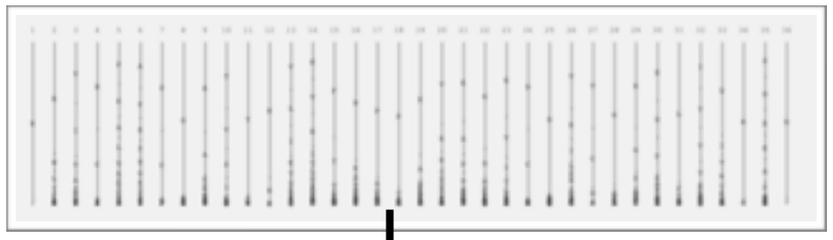
- describe 3 proto-personae

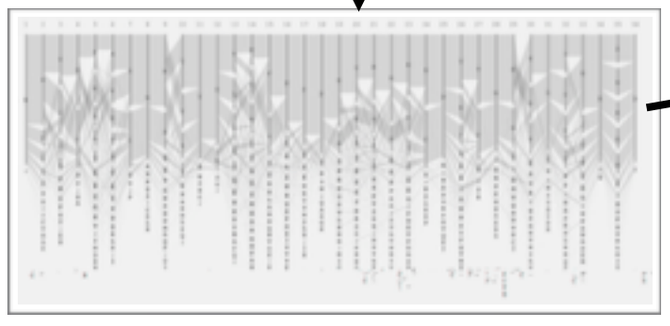
- think of at least 4 goals, and write down goal statements

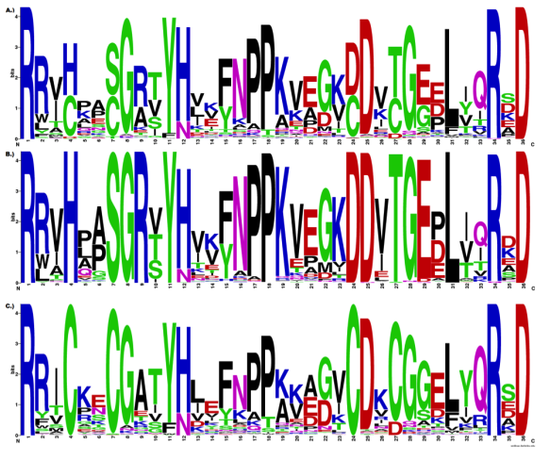
The design

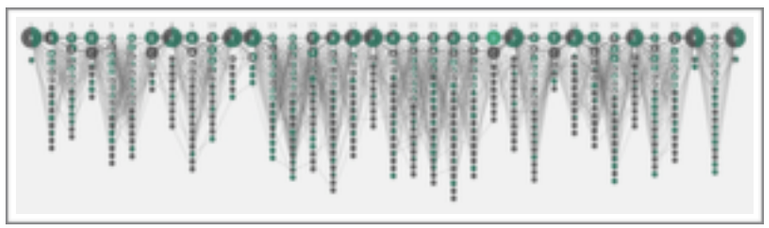


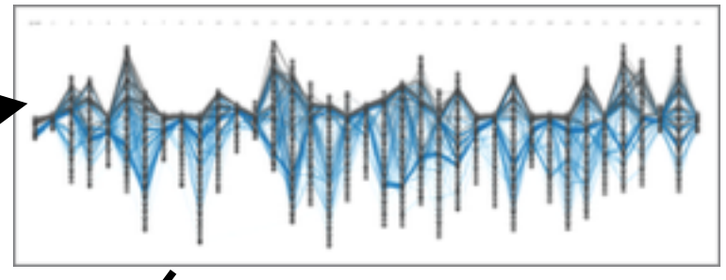


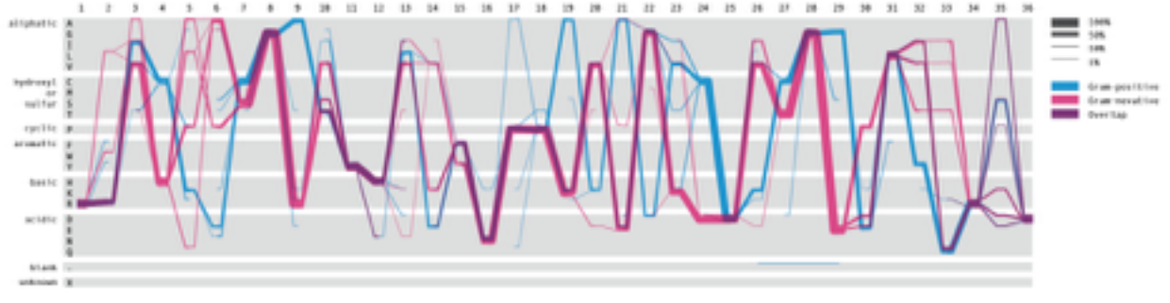


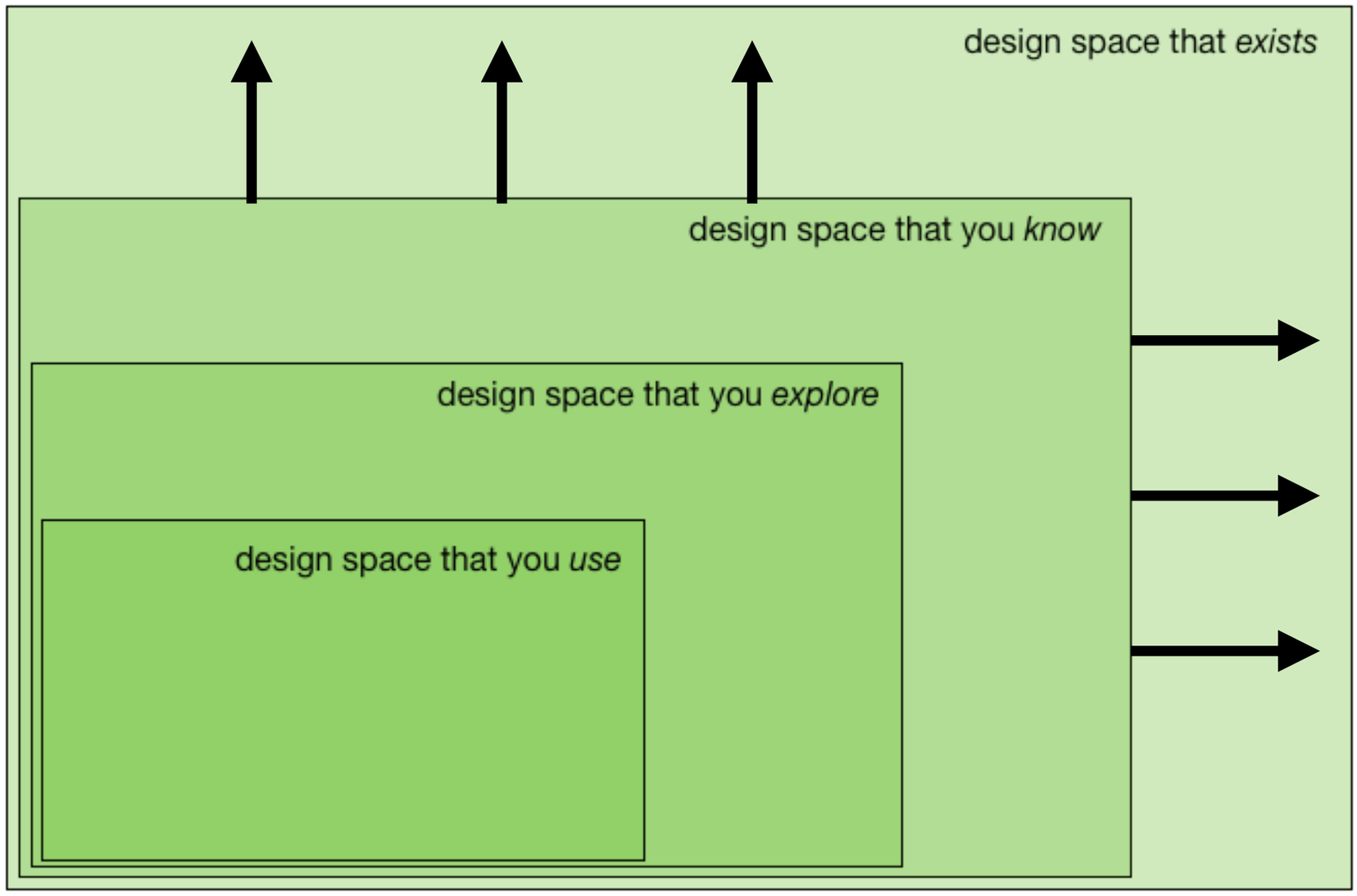









aliphatic
hydroxyl or sulfur
cyclic
aromatic
basic
acidic
blank
unknown
Gram-positive
Gram-negative
Overlap



problem: initial design space that you know is small => how to expand?

Generating ideas - exploring design space

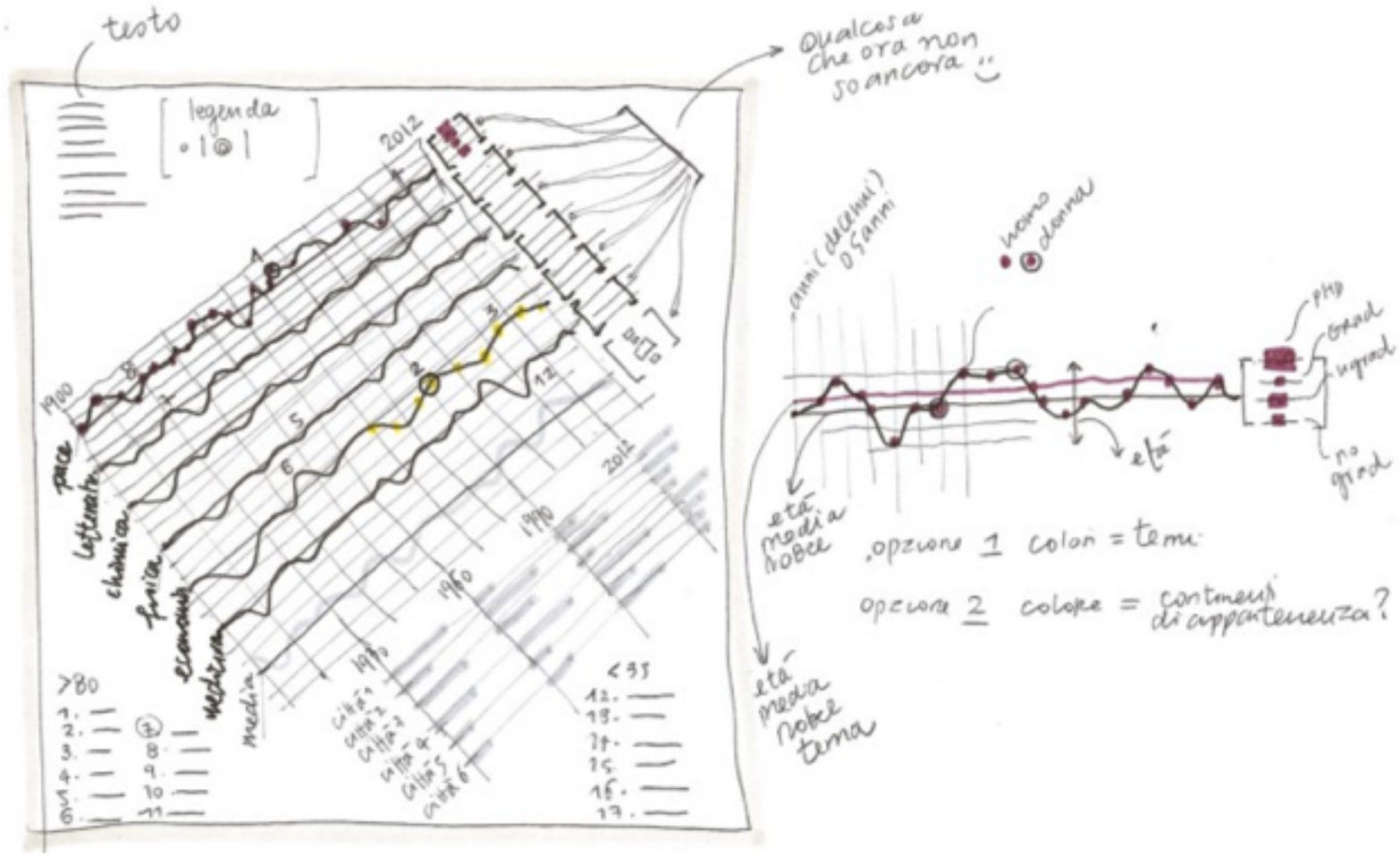
= “ideation”

- use pen & paper!
- approaches:
  - expand your visual library
  - anti-solutions
  - five-design sheets
  - …

Pen & paper

https://t.co/Rvoc8sUOOP

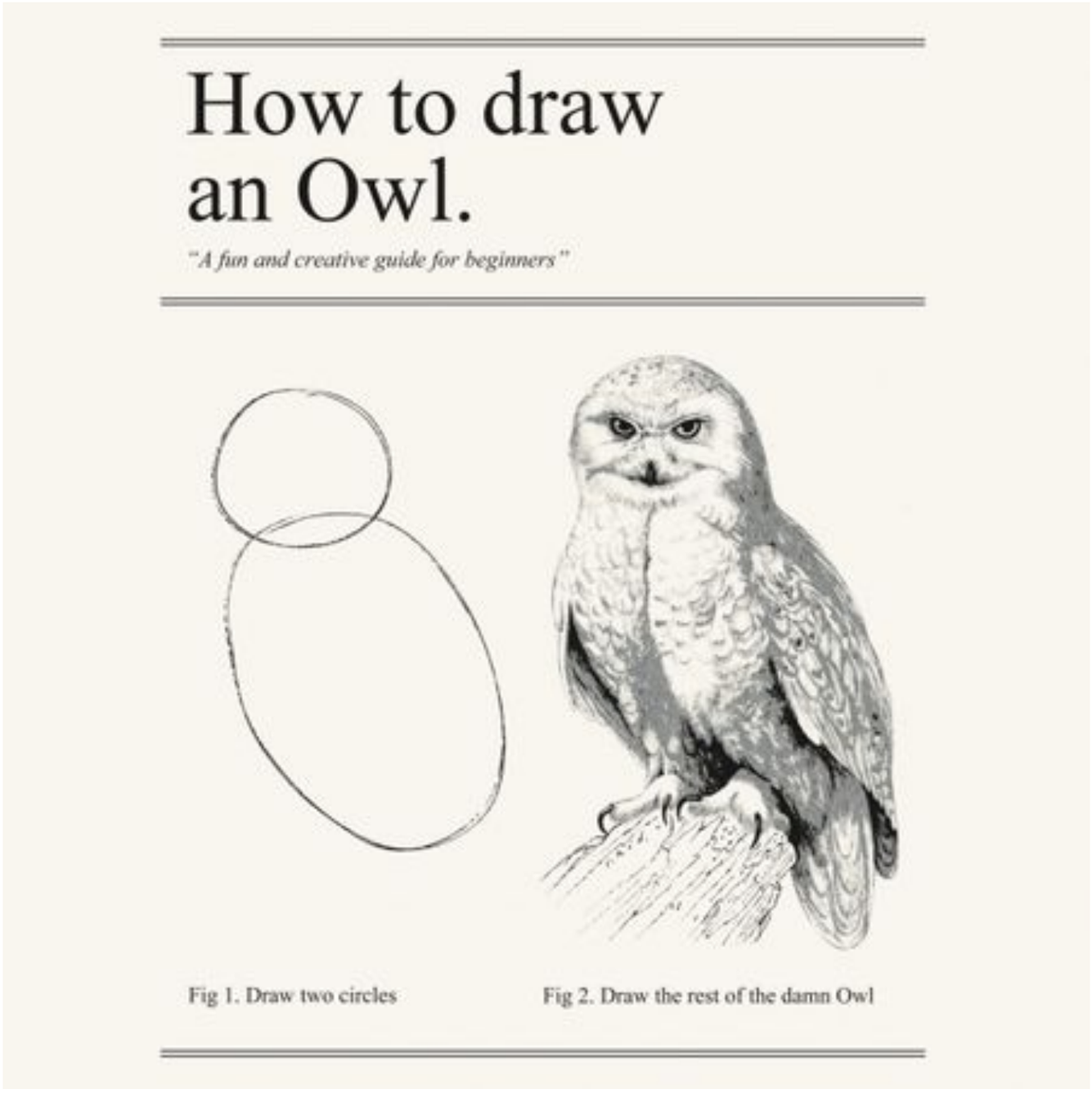
Read this



“Get the big things right during low-fidelity, and the little things will follow in future iterations” (Marc Rettig)

=> biggest benefit of sketching with pen: **sketches are cheap**. (meaning: they are easy and quick to make, so that you won’t cling to them and feel uncomfortable when you discard them)

Intermezzo - “But I can’t draw…”



How to draw an Owl.

“A fun and creative guide for beginners”

Fig 1. Draw two circles

Fig 2. Draw the rest of the damn Owl

[Activity]

- Take a line for a walk
- Draw 8 objects & concepts (10 sec each)

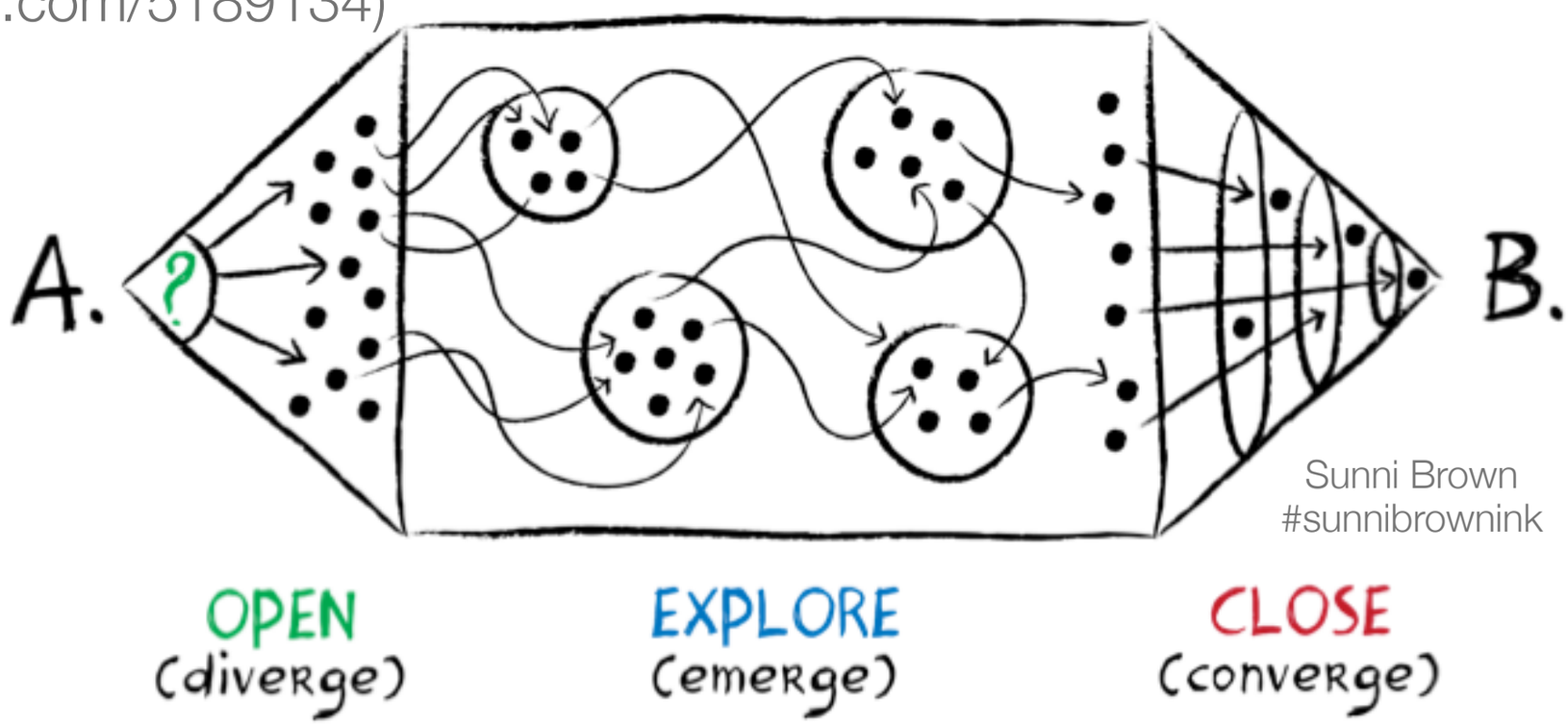
Diverge - emerge - converge

"The best way to have a good idea is to have many" (Linus Pauling)

- Don't arrive with just one idea => if someone critiques that idea, it feels like they critique you

- Arrive with many ideas and *don't commit yourself* to any of them => you can have open discussions

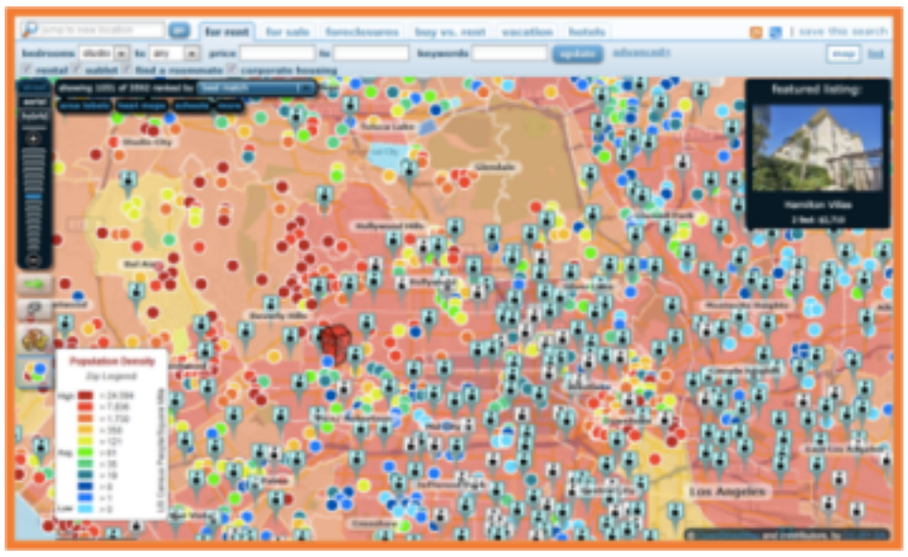
(see Bill Buxton https://vimeo.com/5189134)

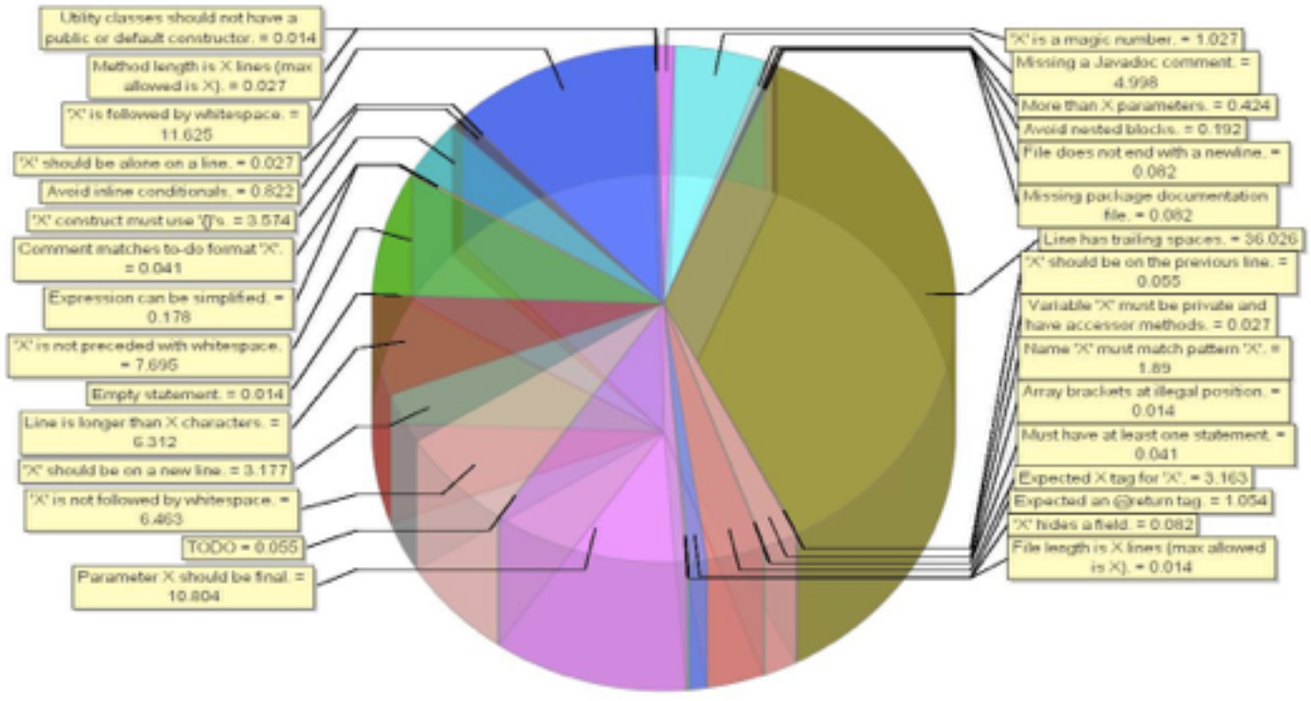


Anti-solution

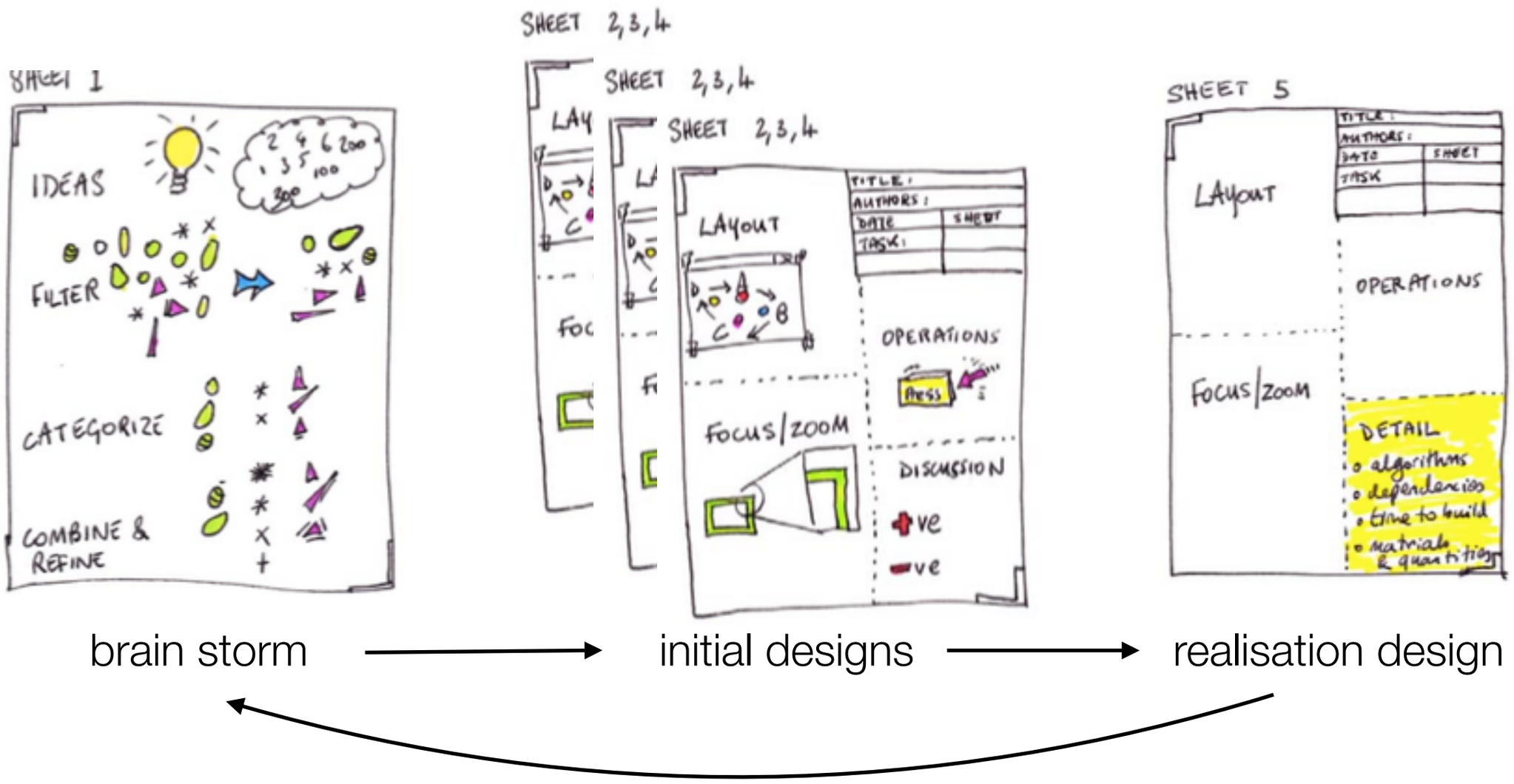
Sketch a collection of designs that are particularly bad at supporting the goals of the (proto-)users (e.g. using “bad” selection of encoding; see Mackinlay)

=> helps to identify what does not work => adds limits to your design space





5-design sheet methodology



Read the paper! (http://fds.design)

5dS: sheet 1 - ideation

- only part that may be >1 sheet
- approaches:
  - **re-work** existing visuals
  - **provoke**: think of impossible solutions
  - glue **dissimilar ideas** together (e.g. use a network to visualize geo-spatial data)
  - **reverse/flip/invert** an idea (e.g. biofabric)
  - …

(see paper for additional approaches)

5dS: sheet 1 - ideation

- parts of the sheet:

1. **ideate** - sketch as many ideas as possible (half-baked, throw a wide net => don’t critique)

2. **filter** - remove duplicate and irrelevant ideas by annotating the existing ideas

3. **categorize** - can these ideas be clustered in some way?

4. **combine & refine** - which visualisations can complement each other?

5. **question** - reflect on what has been created

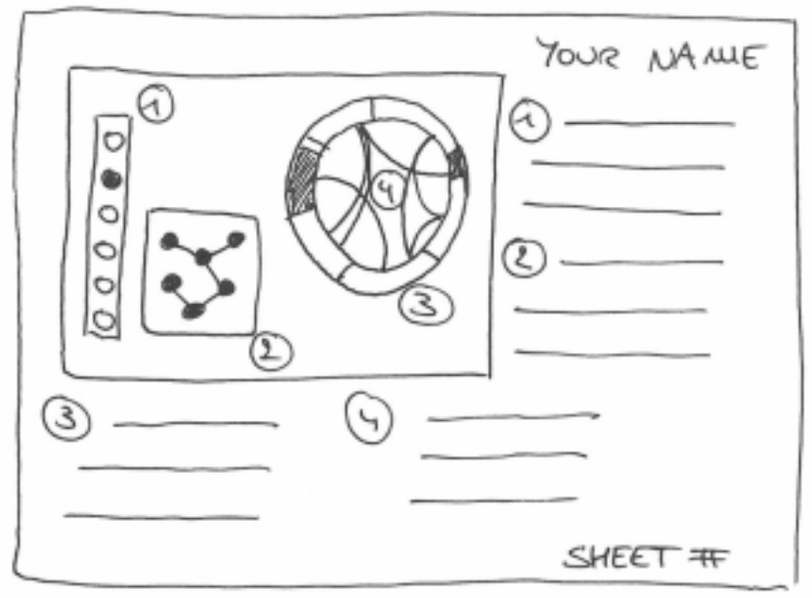
[Activity] 8+8 sketch (individually)

- Fold A3 paper 3 times in half => 8 sections
- Assign goal statements to students:
    - Generate 8 different designs as creative and diverse as possible; *meaningfully* distinct, not just cosmetically (max 20 min total)
    - Choose 2 or 3
    - Generate 8 detailed versions or variations (max 20 min total)
    - Present

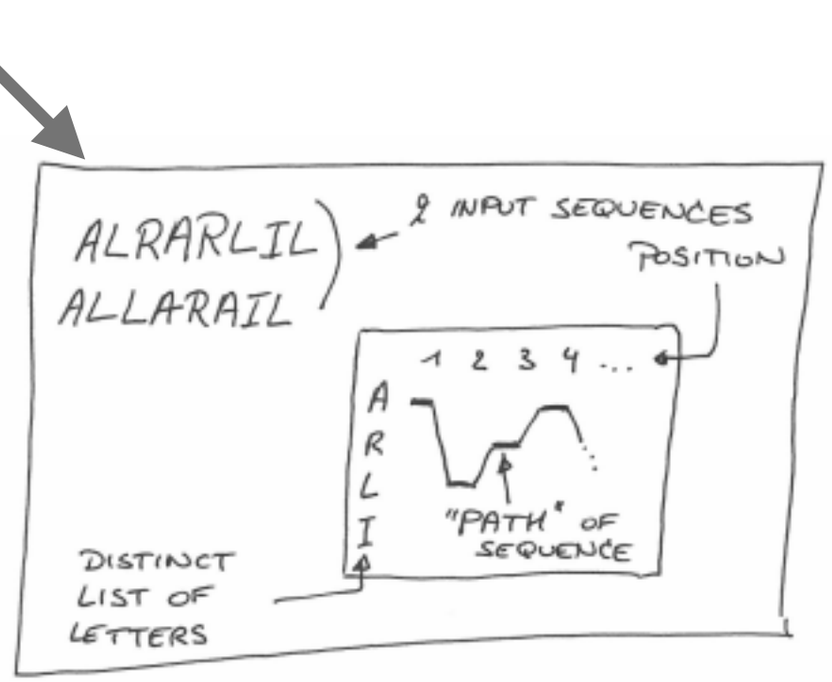
[Activity] Create 5dS sheet 1

- individually; assign statements a-b-c-d-a-b-c-d-a-b-c-d
- use what you already did in 8+8 exercise
- max 15 min

5dS: sheets 2-4



1. **meta-information** - title, name, date, number, …
2. **layout** - overview, =~ sketch of a screenshot
3. **operations** - what interactions are possible? (maybe defer to storyboard; see later)
4. **focus/parti** - explain the central idea; e.g. zoomed into a single component, or a flow diagram
5. **discussion** - advantages and disadvantages



[Activity] Create 5dS sheets 2-4

- max 20 min

5dS: sheet 5

similar to sheets 2-4:

1. **meta-information**

2. **layout**

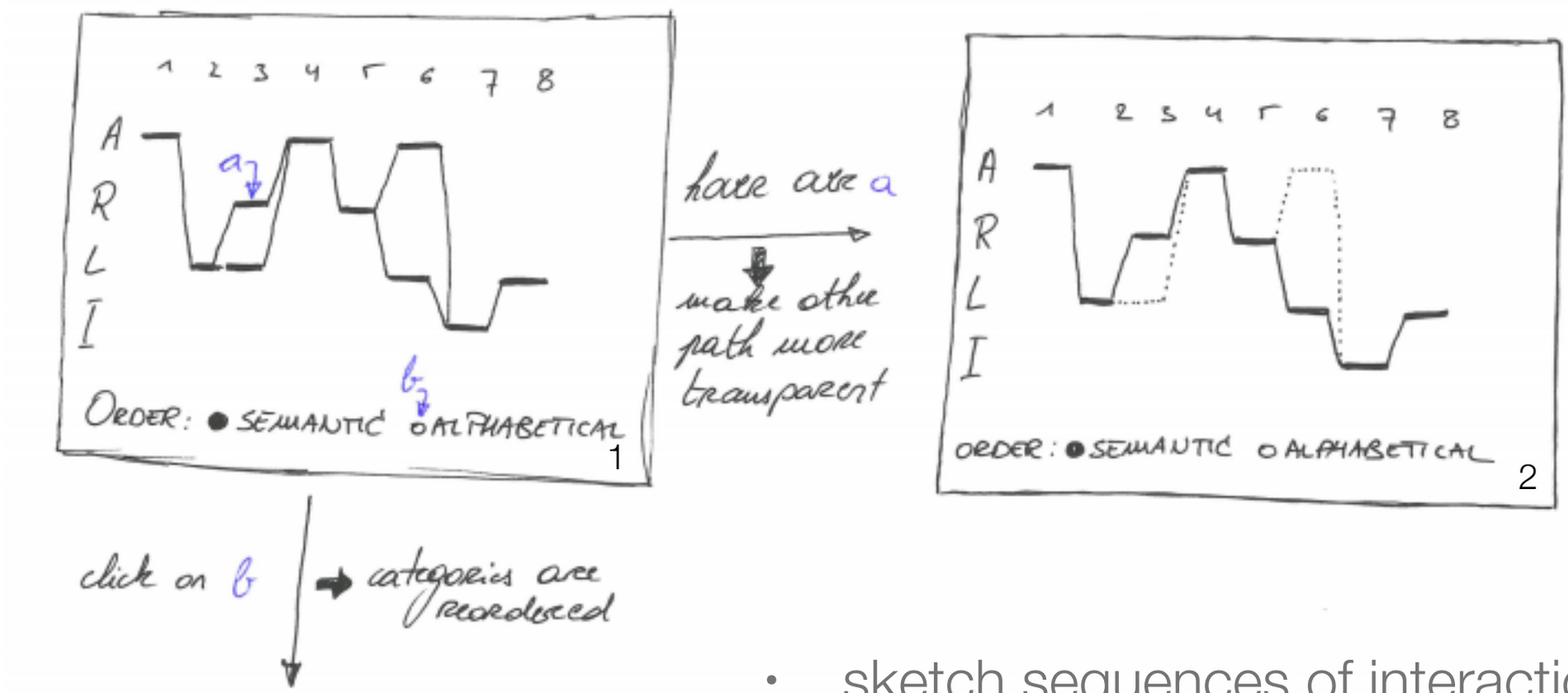
3. **operations**

4. **focus/parti**

5. **details** - algorithms, design patterns, data structures

[Activity] Create 5dS sheet 5

Have interaction? => draw a storyboard



- sketch sequences of interactions
- think about each step and about the transitions
- number each “slide” => refer to detail slides

Critique

- critique != criticism
- method: 2+2
    - what are 2 things to definitely keep
    - what are 2 things that should be changed