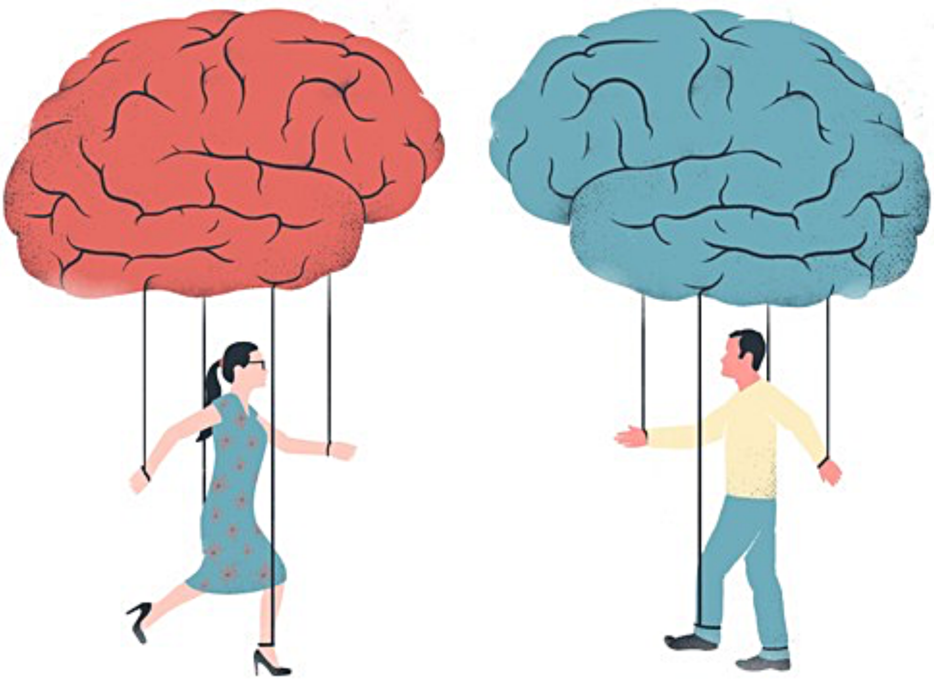
Head First: Data Analysis

01: Introduction to data analysis

Mental models



> *Always make your mental modals as explicit as possible*

Give your mental models the same **serious**, **careful** treatment as **data**.

Question everything

> *You're attracted to what compels you [paraphrased]. (12 Rules for Life, Jordan Peterson, 2018)*

So, what you focus on *matters*: the world will represent to you what you believe. Be careful of blindly following one perspective or mode of thought.

Your mental model will reveal the features of the world you focus on. Take, for example:

1. 🍸 Hipster place
2. ☕ A regular place

I perceive a hipster place as *superficial* and *exclusive* — that doesn't make it true.

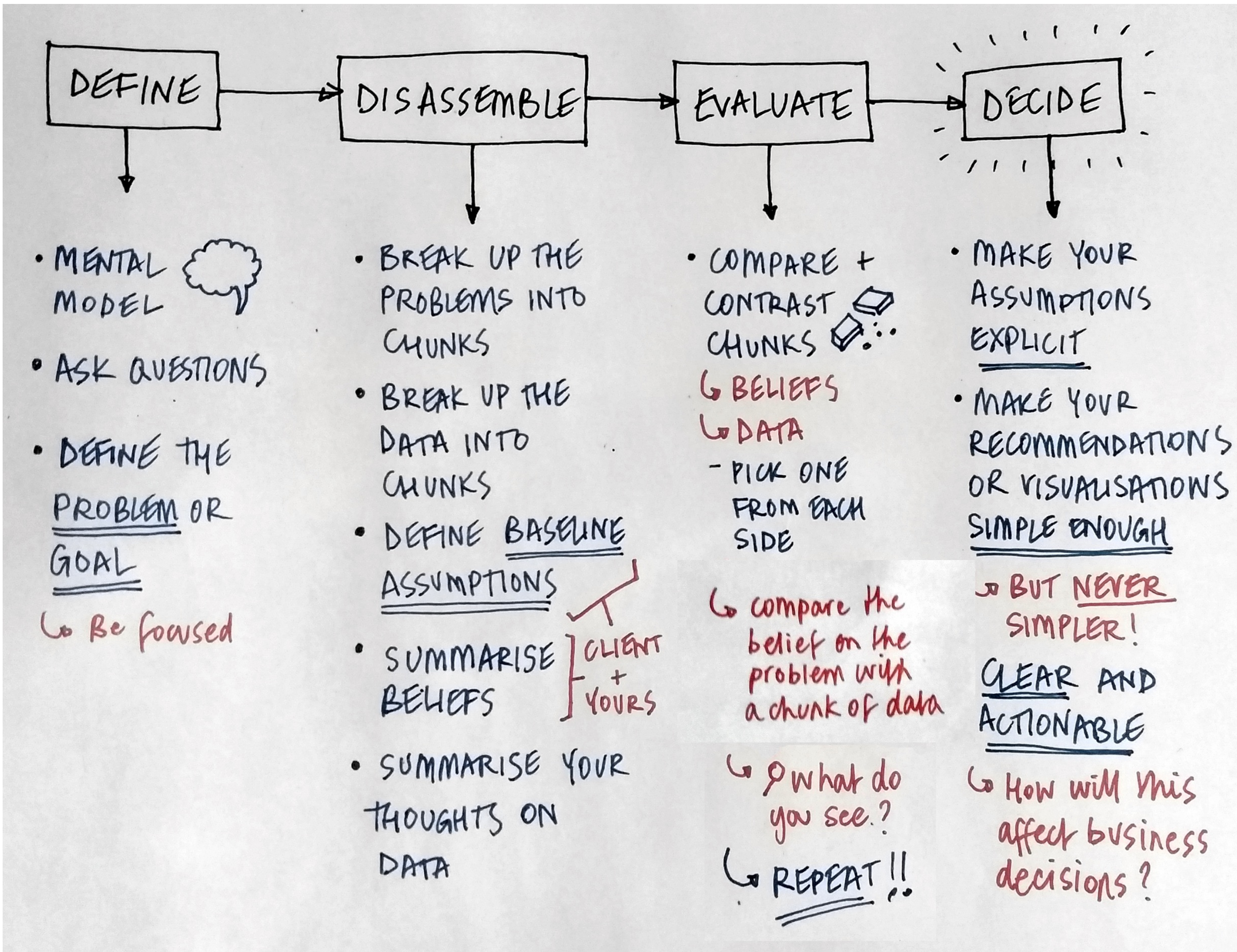
- *Meaning* is what draws you in.
- Focus is what has *meaning* for you.

Your mental model

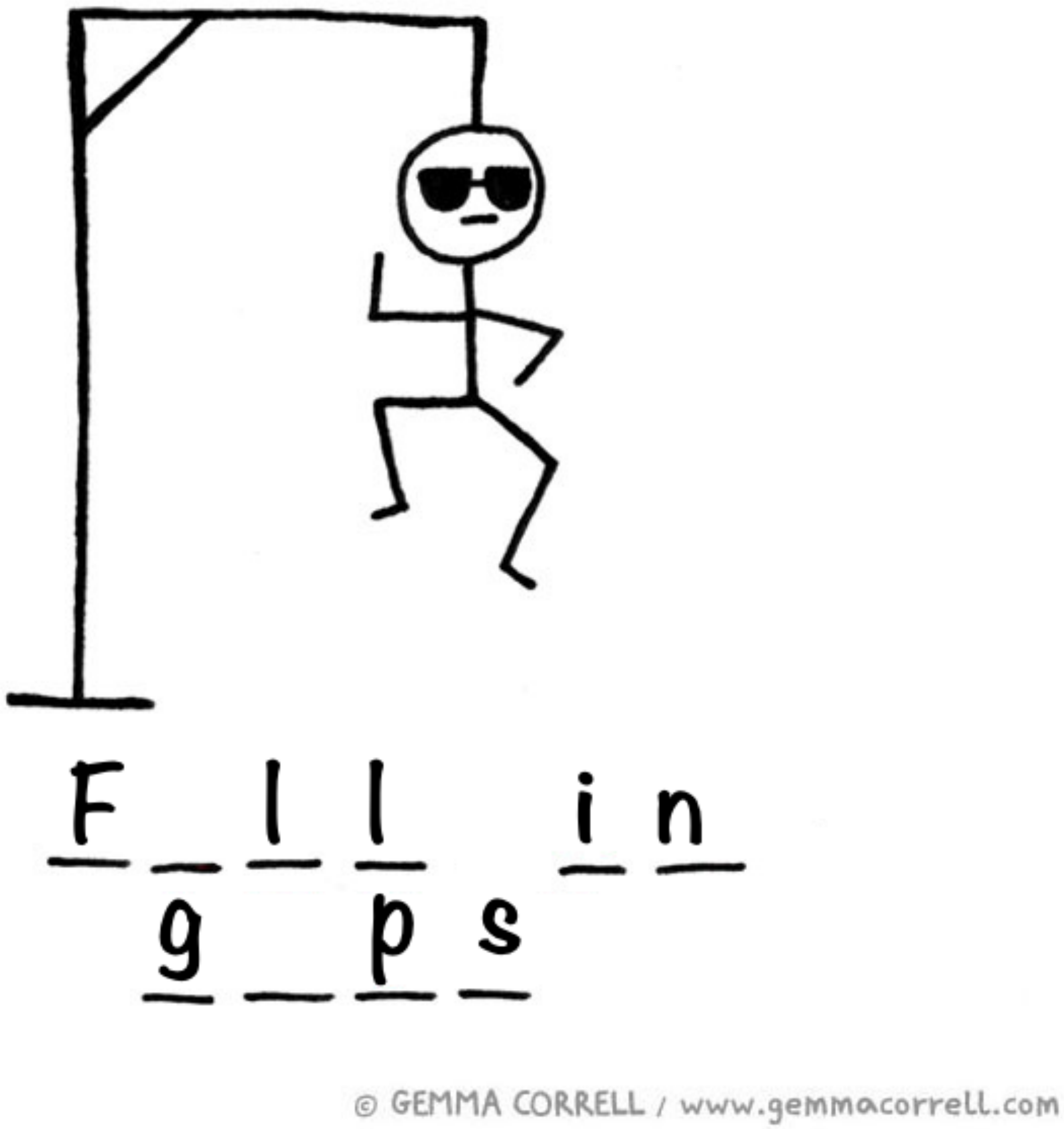
> *You have a mental model — they do too! Be concious of that.*

- **Keep this in mind** and you'll see what's important
- **Your statistical model** is dependant on your on your mental model
  - To pick the right statistical model
  - You need be aware of your mental model
- **You will fail** in your analysis (or recommendations) if your mental model is wrong
- **Your questions** or theories change, depending on your mental model

Define, disassemble, evaluate, decide



Filling in the gaps



A mental model helps you *fill in the gaps* in your knowledge.

- Find problems and gaps in knowledge
- Test these with data wherever possible

Make links between your data and knowledge — your mental model will help you understand the world, filling in the gaps as you go.

> *Always include what you don't know, as well as what you do*

- What do you know?
- What don't you know?
- What don't you know, you don't know?

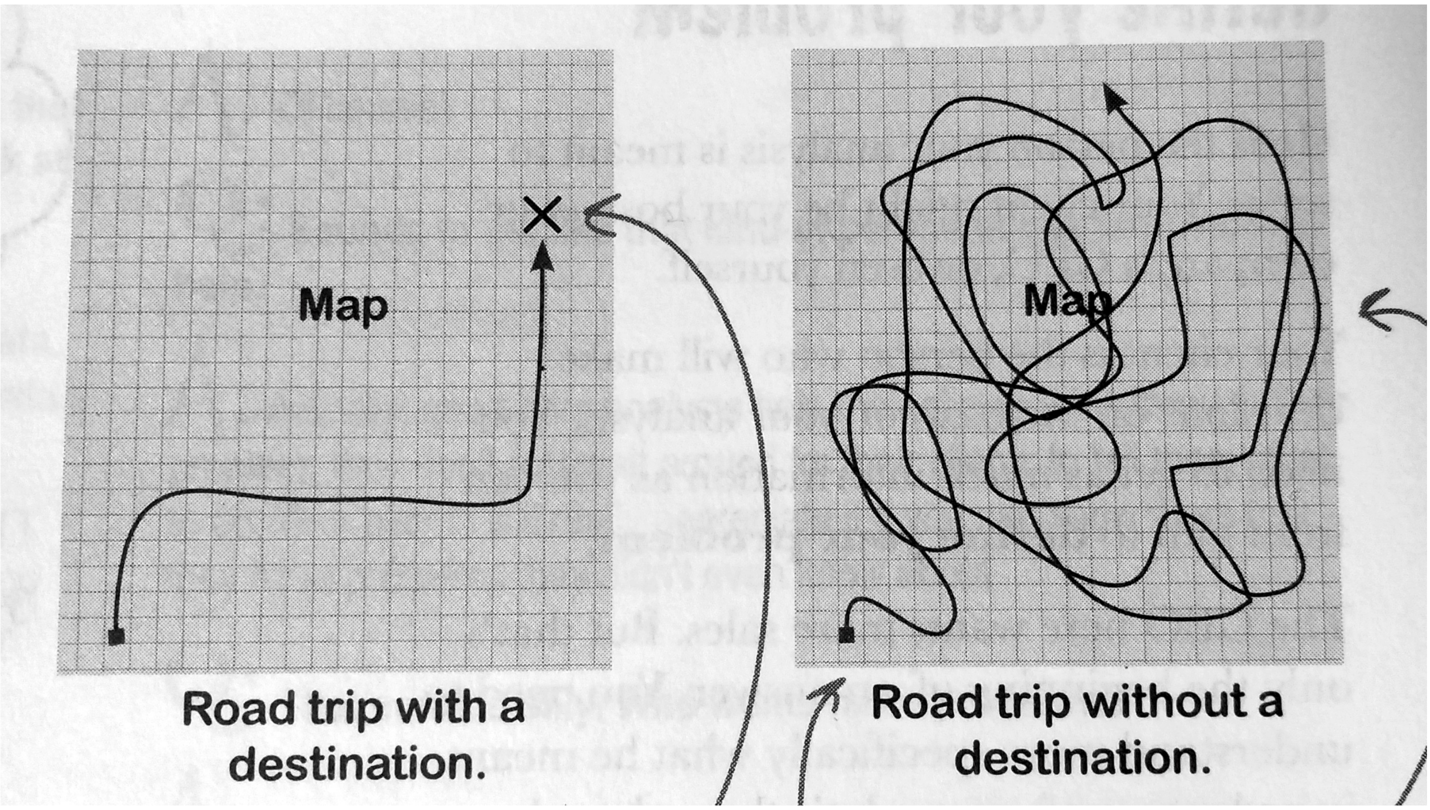
It's ok to be wrong. you'll make mistakes so just keep asking the question: does it work?

Your baseline assumptions

> *Todays research become assumptions you take into tomorrows research*

- **Make your assumptions explicit**
    - so you can recognise errors …
    - then correct them
- **You can't test everything**
    - but everything should be testable

FOCUS



Your mental model and baseline assumptions help you find your focus

Stay focused on what you're trying to achieve, no matter how much data you have at your disposal.

> *BUILD → MEASURE → LEARN*

1. Your mental model
2. Your baseline assumptions

3. Together, become your focus 🔦

It’s an ongoing process

Data analysis is an ongoing process. You can’t know everything and todays success could be tomorrows failure.

1. **Add new data** when it becomes available
2. **Reassess your mental model** often:
    - Your mental model influences your data model
    - Your data model influences your mental model
3. **Stay focused on your goals** and **make your assumptions explicit**

02: Test your theories

Problems to be solved

What, exactly is the problem?

There’s a few ways to gauge the problem:

1. Client interviews
2. Customer interviews
3. Personas or archetypes

Gather your **baseline set of assumptions** about their problems. They won’t necessarily true as these can be *feelings* (not *facts*), but it’s a good place to start:

1. Define the problem
2. Write your client (or customer) story
    - Tasks they want to perform
    - Assumptions they might have
3. Write down their *baseline set of assumptions*

Finding the TRUTH™

> *TRUTH™ is somewhere between **what they say** and **what they do***

Let's take the example of a meal out. Jim is asked by the waiter "was everything ok for you?"

- Does he answer honestly?
- Or does he say what they want to hear?

Perceptions are not always the same as *objective* behaviour. Conversations help us empathise through *speech* and *feeling* (empathy) — but always check your facts:

- **Qualitative** *gathers* insights
- **Quantitative** *validates* insights (always test theories)

It takes a little of both!

Comparing Data

It's best to break your data into chunks

Data is **only** interesting when in **comparison** to other data. What changes or patterns do you see?

> *Always use the method of comparison in analysis*

- *November* compared to *December*
- One statistic (or data point) compared to another

Speaking your TRUTH™

> ⚠️ *Always make comparisons explicit*

1. Clearly state your *assumption* about the *comparison*
2. Clearly communicate this to client and yourself

Observational Data

Qualitative doesn’t tell the whole story, but gathers insights for further experimentation:

1. Watch. Observe.
2. See what they do.
3. See what groups they assign themselves to.
4. Take inventory (track observations)

Flipping the theory

Sometimes you'll have to flip the theory to see if the opposite is going on. A good example of this is **value perception**:

- Is it a customer sense of value for their money?
  - "The price is too high for what you're giving me!"
- Or does a customer value something more when the price is higher?
  - "The price is too cheap! It mustn't be good quality!"

⚠️ Beware of first impressions

Observational data has it's limitations

- Data can sometimes lead you down the wrong path …
- Be careful you're not drawing the wrong conclusions
  - Take observational data with a pinch of salt

Say it in pictures

Draw pictures of how you think things relate. This helps you make your ideas explicit and easy to view at-a-glance.

Brownie sales image

Confounders

> ***Always** consider how **confounding** may **affect results***

- Confounders should make sense in the context of your analysis
- A confounder in one observational study, may not be relevant in another

Imagine you're conducting two studies. The economy is shot, consumer spending down, and your stores are in both *rich* and *poor* areas — sales are down and you're trying to find out why. For each theory, what are the confounders?

☕ Value perception	☕ Temperature
📍 location	~~📍 location~~
-	👨‍🍳 staff member
-	⏰ time of day

Managing Confounders

These smaller chunks are more **homogenous**. In other words, they don't have the internal variation that might skew your results and give you the wrong ideas.

Here is the Starbuzz survey data once again, this time with tables to represent other regions.

Here's the original data summary.

Starbuzz Coffee: All stores

Summary of marketing surveys for six months ending January 2009. The figures represents the average score given to each statement by survey respondents from participating stores.

	Aug-08	Sept-08	Oct-08	Nov-08	Dec-08	Jan-09
Location convenience	4.7	4.6	4.7	4.2	4.8	4.2
Coffee temperature	4.9	4.9	4.7	4.9	4.7	4.9
Courteous employees	3.6	4.1	4.2	3.9	3.5	4.6
Coffee value	4.3	3.9	3.7	3.5	3.0	2.1
Preferred destination	3.9	4.2	3.7	4.3	4.3	3.9

Mid-Atlantic stores only

	Aug-08	Sept-08	Oct-08	Nov-08	Dec-08	Jan-09
Location convenience	4.9	4.5	4.5	4.1	4.9	4.0
Coffee temperature	4.9	5.0	4.5	4.9	4.5	4.8
Courteous employees	3.5	3.9	4.0	4.0	3.3	4.5
Coffee value	4.0	3.5	2.9	2.6	2.2	0.8
Preferred destination	4.0	4.0	3.8	4.5	4.2	4.1

Seattle stores only

	Aug-08	Sept-08	Oct-08	Nov-08	Dec-08	Jan-09
Location convenience	4.8	4.5	4.8	4.4	5.0	4.1
Coffee temperature	4.7	4.7	4.8	5.1	4.5	4.9
Courteous employees	3.4	3.9	4.4	4.0	3.5	4.8
Coffee value	4.3	3.8	3.2	2.6	2.1	0.6
Preferred destination	3.9	4.0	3.8	4.4	4.3	3.8

SoHo stores only

	Aug-08	Sept-08	Oct-08	Nov-08	Dec-08	Jan-09
Location convenience	4.8	4.8	4.8	4.4	4.8	4.0
Coffee temperature	4.8	5.0	4.6	4.9	4.8	5.0
Courteous employees	3.7	4.1	4.4	3.7	3.3	4.8
Coffee value	4.9	4.8	4.8	4.9	4.9	4.8
Preferred destination	3.8	4.2	3.8	4.2	4.1	4.0

These groups internally homogenous

Breaking the store survey results by region

Break data into chunks

Sometimes it's a good idea to split out the data into smaller chunks, where the data is less likely to be biased. It'll help you manage confounders, some of which could be linked:

- 🔗 location
- 🔗 wealth
- age
- weather
- …

Are your observations valid?

> *If I train staff better, will the temperature of the coffee improve?*

- Sometimes a theory will overlap with observational data.
- Sometimes observations will give rise to a theory.
- Observational data is often unlikely to predict the future

You'll need an experiment if

- Your observational data doesn't describe this theory
- There's not enough data to validate a theory, or give you predictions
- Your observational data isn't strong enough

If any of these are true, it's **just a theory**!

Experimenting

> *You might have two or more theories: You need an experiment!*

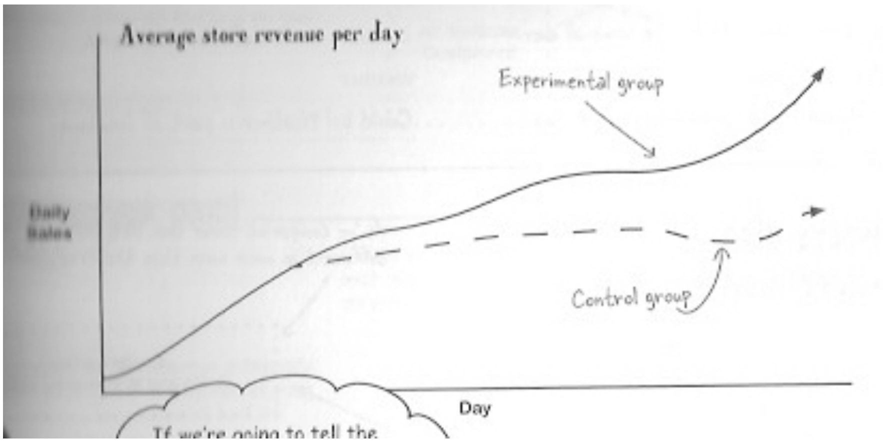
You'd like to find out why people aren't buying.

1. Mark thinks it's a *pricing* problem (discount, reduce price)

2. Ben thinks it's a *perception* problem (rebrand)

Who's right?

It's important to set a baseline



Let's take Mark's theory. If we went ahead and changed price across all stores, we can never be quite sure if our changes were the *real* reason for failure or success!

- **Set a baseline control group**
    - An A/B test or experiment
    - A multivariate test
- **What would've happened *without* the experiment?**
    - Did it have the desired effect?
    - Have you ruled out any confounders?

You can't *always* set a control, but always aim to.

Randomise

> *Random selection gets you as close as possible to causal relationships*

We've already tried:

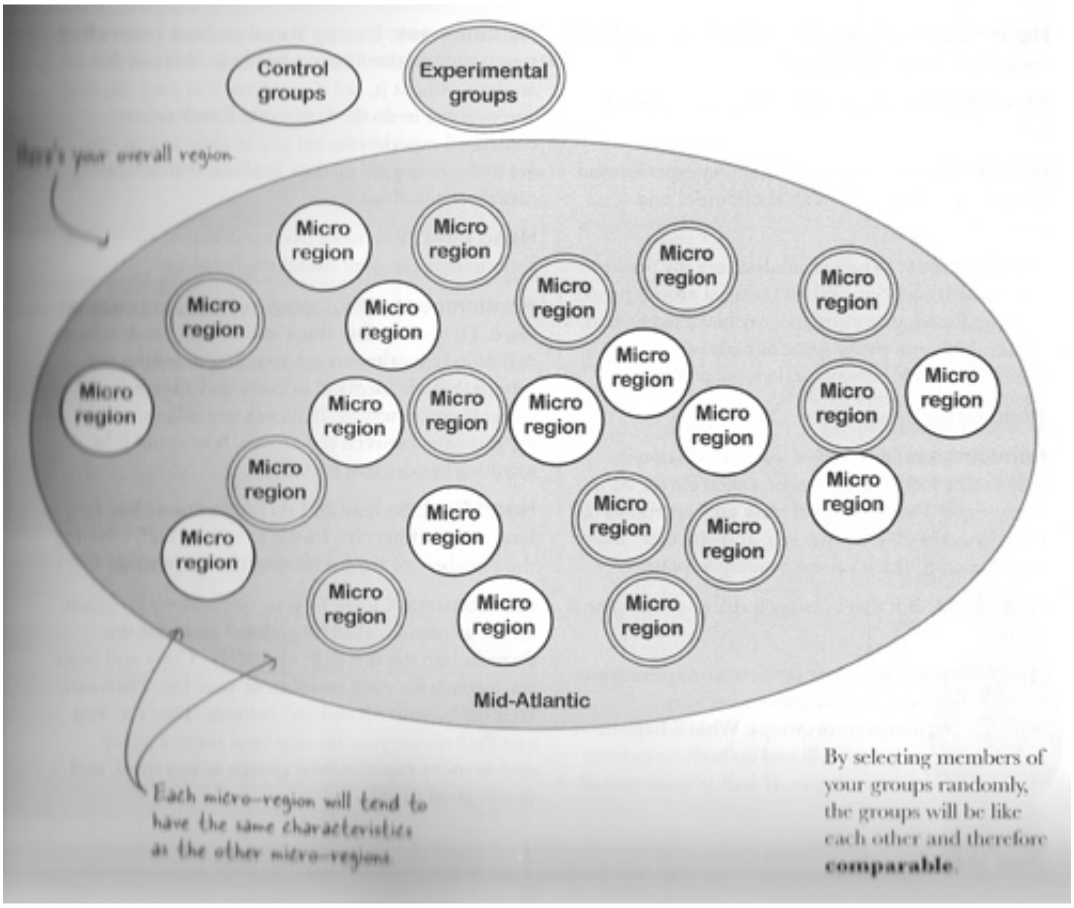
- Chunking the regions
- Running an A/B experiment

Confounders may still plague our experiments! In order to reduce our confounders *our groups need to be the same*. We could try the following:

1. ~~Charge every other customer a different price~~ (wouldn't go down well)

2. ~~Historical controls (has historical confounders)~~
3. ~~Randomly assign stores control and experiment groups~~ (customers may choose cheapest)
4. Divide big geographic regions into micro regions (and randomly assign to control/experiment)

When you randomise, the factors that might otherwise become confounders get equal representation in control/experiment groups. So if we had a hidden confounder “X”, both groups should contain the hidden confounder in (roughly) equal amounts. It *should* (but not guaranteed to) affect your groups in equal ways.



Randomise your experimental groups to minimise confounders

So, here were the steps we took:

1. First try to avoid any obvious confounders, like location
    - See Starbuzz example [**pg: 67** grouping by micro-regions]
2. Next, **randomly assign** those groups to the **control** and **experiment**
    - Ideally you’d equal out everything *except* the variable you’re testing for, but that’s best case scenario.

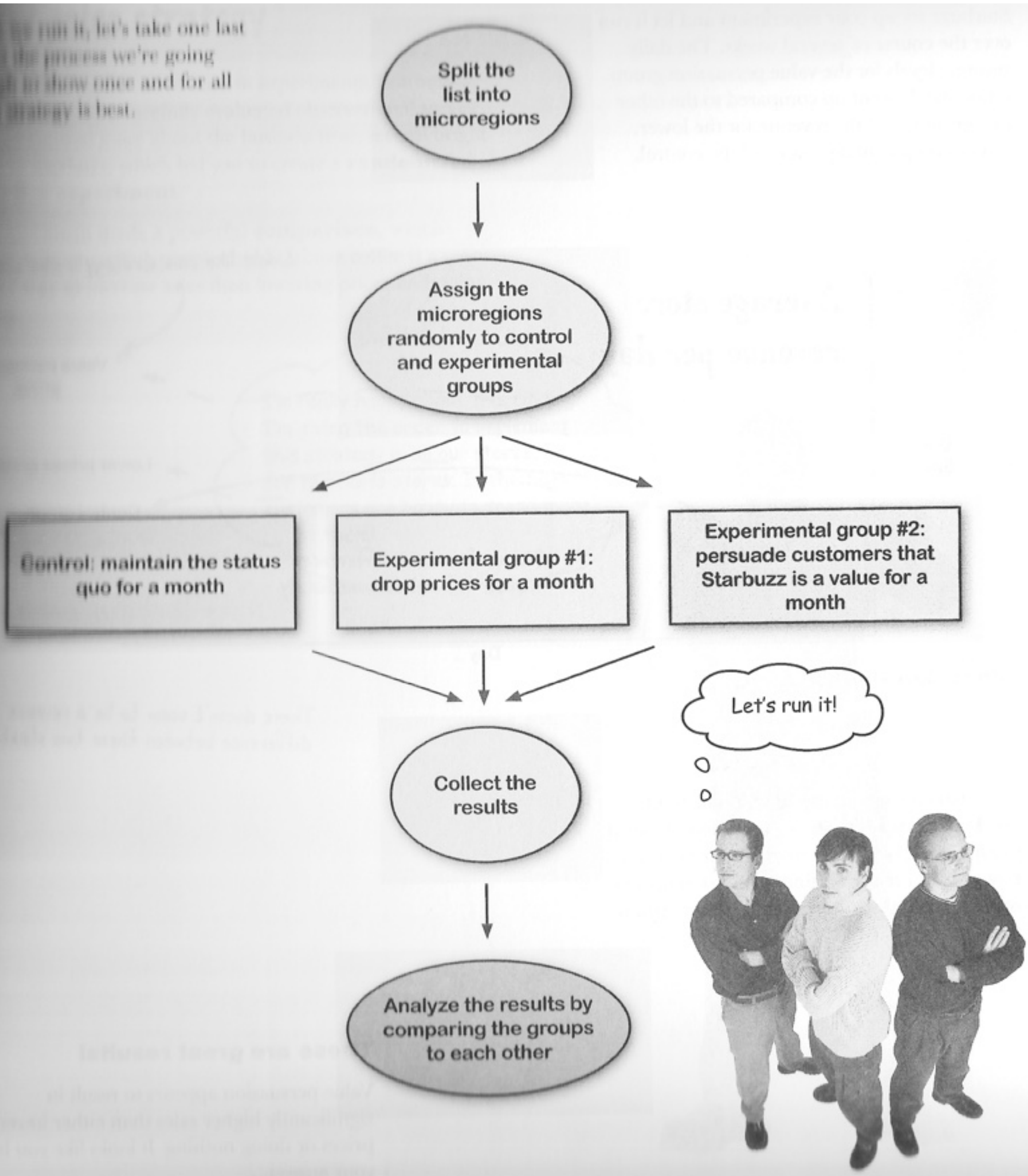
⚠️ Other confounders to consider

- Sample size
- Observer or subject bias

Experiment flow

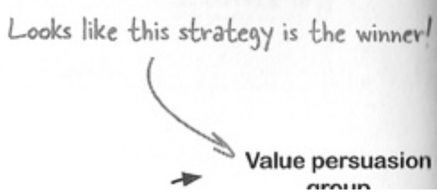
Theories, randomisation, experiment

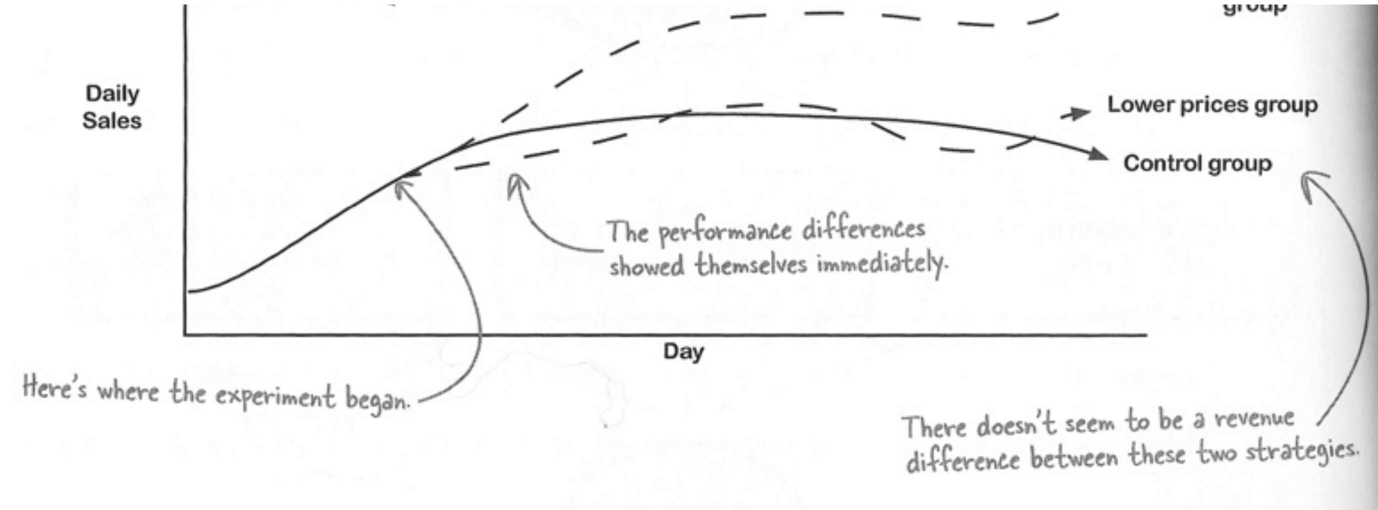
...e run it, let's take one last
... the process we're going
... to show once and for all
... strategy is best.



Starbuzz set up your experiment and let it run over the course of several weeks. The daily revenue levels for the value persuasion group immediately went up compared to the other two groups, and the revenue for the lower prices group actually matched the control.

Average store revenue per day





This chart is so useful because it makes an excellent **comparison**. You selected identical groups and gave them separate treatments, so now you can really attribute the differences in revenue from these stores to the factors you're testing.

03: Optimization and control

Elements of control

> *Always ask for this data*

When receiving a brief, it's helpful to collect data and group it by:

- Elements you *can* control
- Elements you *can not* control

Decision variables and constraints

> *The balance of your decision variables and your constraints determine your profit*

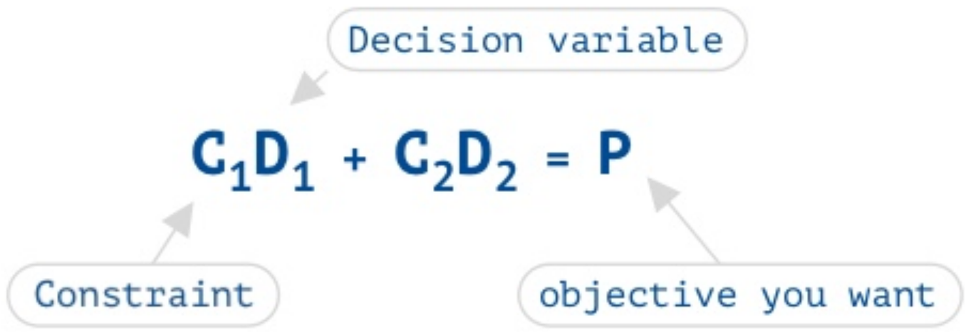
1. Things you can control	2. Things you can't control
Decision variables (can alter profit)	*Constraints (can't alter profit)*
— How much *Brand A* to buy	— Cost of each brand (wholesale)
— How much *Brand B* to buy	— How profitable each brand is
	— How much it costs to deliver
	— How long it takes to deliver

Optimization

> *All optimization problems have constraints and an objective function*

- **You can maximise profit** by choosing the right *product mix*
- **Objective functions** help you decide how many of each product range (for example) to produce.

An objective function



An *objective function* is used when you want to find a `max()` or `min()` of something. This could be profit, manpower, or logistics. You'll combine the things you *can* change with things you *can't* change to reach your objective.

Complex optimisations will need more sophisticated functions, but in essence, this is how an objective function looks.

Example: objective function

> *If I increased production of (☕ A) or (☕ B), which would maximise profit?*

- The *constraints* in this case are profit for each coffee.
- The *decision variable* is how many of each to produce.

```
a = $profit_per_☕ * number_of_☕   # total cup A profit
b = $profit_per_☕ * number_of_☕   # total cup B profit

a + b = profit
```

⚠️ It's rarely that simple!

Constraints are rarely this simple and will need further thought — the above example is a very basic constraint. Other variables might be:

- Time
- Materials
- Multiple products
- etc

Finding the right product mix

Optimization bar chart

We're trying to find the right **product mix** which will `max(profit)`. We could present our objective functions with a bar chart:

Decision variable chart comparison

- **Product mix 1** is good to go!
- ~~**Product mix 2** breaks our constraints~~

Our bar chart is great if we only have one product line or a single constraint, but what if we have more? Which type of chart is will best present our objective functions?

A scatter plot chart is better

Scatter plot chart example

If we plot our product mix on a chart like this, we can easily spot our "good to go", or "feasible" region.

Changing the feasible region

Adding *constraints* to this chart *changes the feasible region* and we can figure out what is the **optimal value**

- What if we only had a certain amount of coffee beans to roast and grind?

- These coffee beans must be split between ☕️ `A` and ☕️ `B`

Diagram showing feasible regions of coffee cups

- We could make `500 * ☕️ A` or `400 * ☕️ B`
- Or, we could split (mix) the coffee beans between them

Because of these new constraints, we are even more limited in our product mix options. We can now use our objective function to work out the `max(profit)`

Image of A, B, C versions of product mix

```
A == ☹
B == 😌   # ($5 * 100) + ($4 * 200)= $1300
C == 🤑   # ($5 * 50) + ($4 * 300) = $1450
```

Product mix `C` brings the highest profit of the three!

Optimisation is useless …

… *unless* we make the right assumptions

In our previous models, we're picturing an ideal world where customers will buy whatever we make. This never happens! (or at least, it's incredibly rare).

😏 Customer assumptions. What will they buy?

> ***All models are wrong. But some are useful!*** *[In other words, reality is complex — provide the closest, most useful model you can] (George Box)*

IDEAL vs ACTUAL image

Will the sky fall down? Or might you just lose some cash? How closely your assumptions should mimic reality *depends on how important your results (or analysis) are*.

Consider what assumptions you need to mimic reality

1: What are the variables?

Constraints	Decision variables
1. How long to produce	1. How much of each to produce
2. Cost of production	
3. Profit	

… plus

- real world constraints
- Buyer behaviour
- location

2: What's the objective function?

$$(C_1D_1) + (C_2D_2) = \boldsymbol{P}$$

… plus

- C_3: what sells?
- C_4: in which store?
- C_5: is there any industry data available?

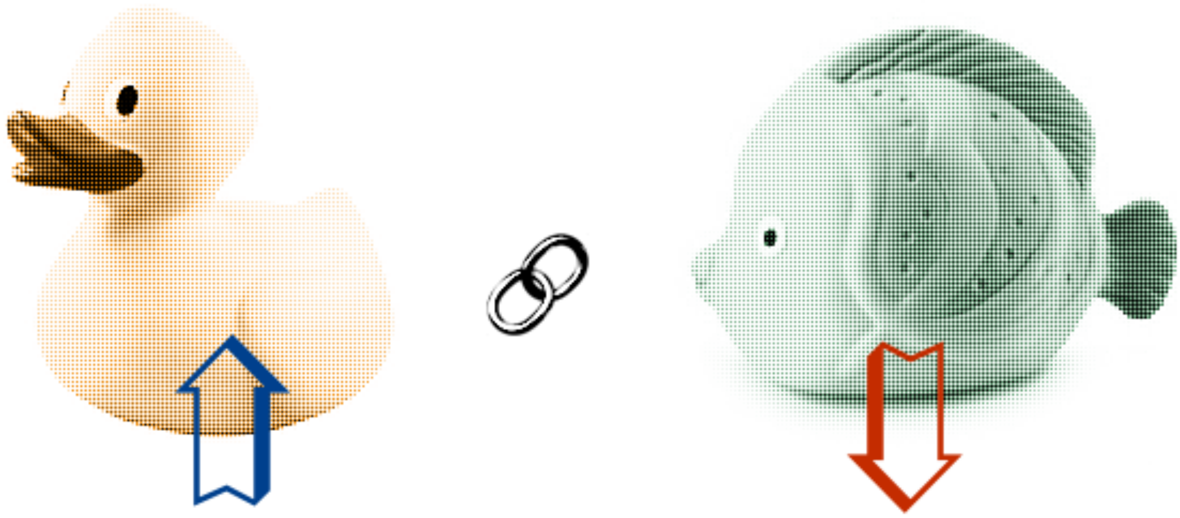
3: Ask yourself …

1. What do you want to achieve?
2. What are the constraints?
    1. What's stopping you, or affecting you from getting it?
    2. **List as many constraints** as possible, that you could represent quantatively with data!

Independent or linked?

basic chart for linked variables

Does the sales data tell you if items are linked or independent?



Does one go up when the other goes down? This is negatively linked

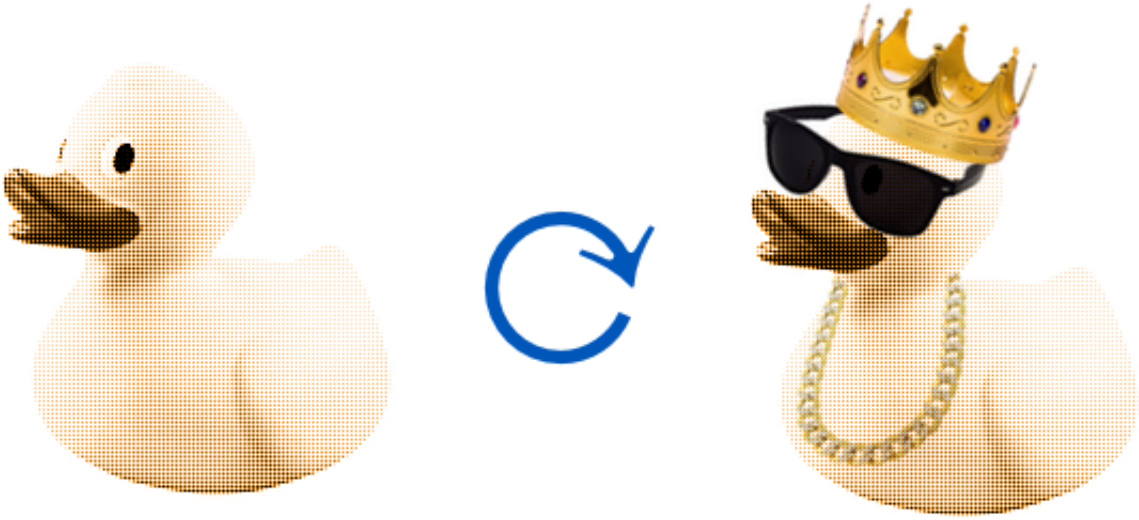
Variables can sometimes be linked. Don't assume two variables are independent of each other. Any time you create a model, take care to *make clear your assumptions* on how variables relate to each other.

If you don't think they relate to each other, make this clear too.

- [ ] What is the relationship as a whole (total sales)
    - [ ] Will sales drop?
- [ ] What relationships does it show?
    - [ ] Which to stock?

The future will change, you can't predict it.

I've got news for you son: all your data is observational, and your model can break at any time. The best you can do is to be clear on your assumptions and mental model (this time), and keep it up-to-date — *learn as you go*.



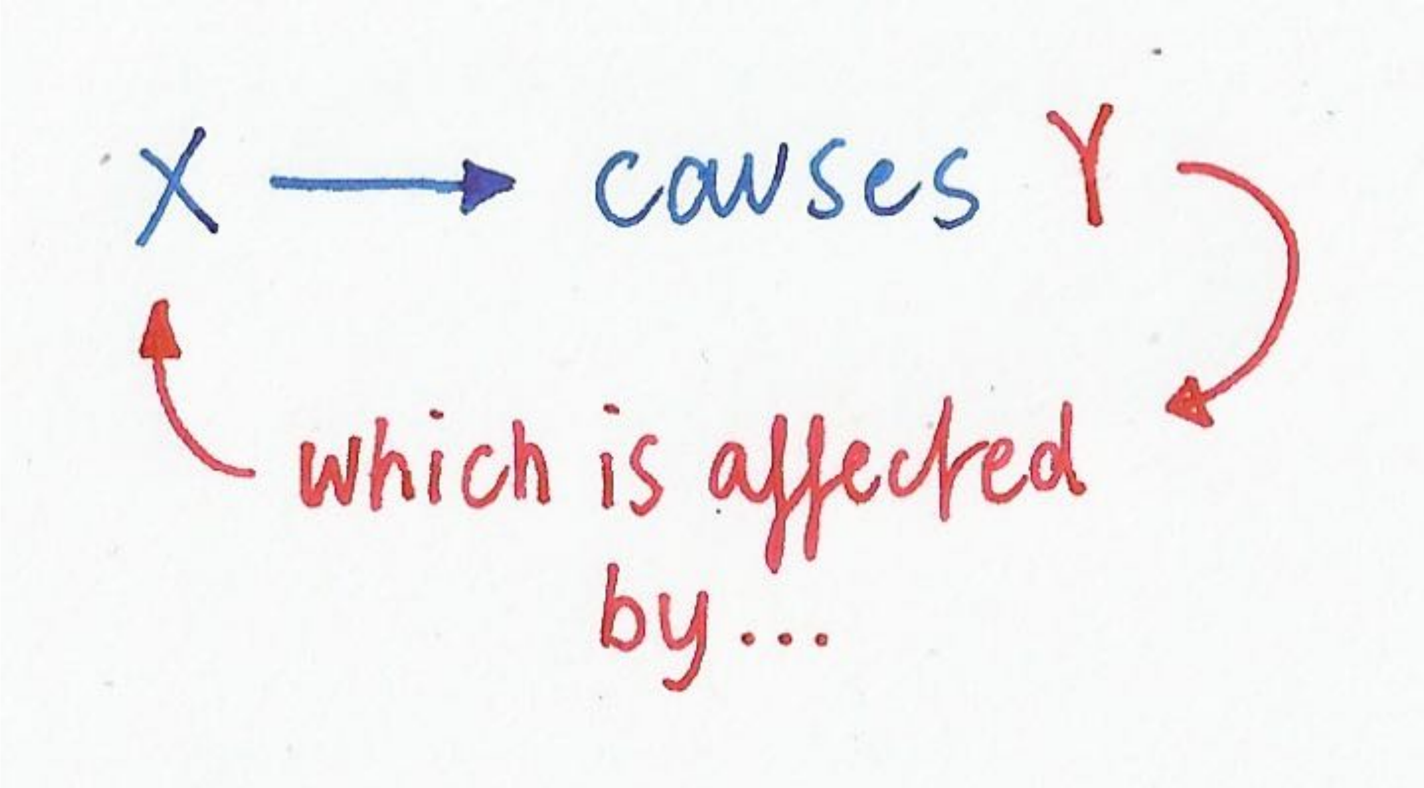
> ***Pimp out your model when things change!*** *If the landscape (or the data) changes, you need to reassess your assumptions and upgrade your model.*

Data analysis is all about *iteration*:

1. How have the relationships changed?
2. How does it compare to last week? Last year?

04: Data visualisation

Scatterplots



A scatterplot explores *causes*. It's an example of *exploratory data analysis* (having a peak at the data that needs testing) where you're asking two main questions:

1. What is your objective or goal?
2. What are the expected causes for this goal to move ↑ or ↓?

Look at two variables together:

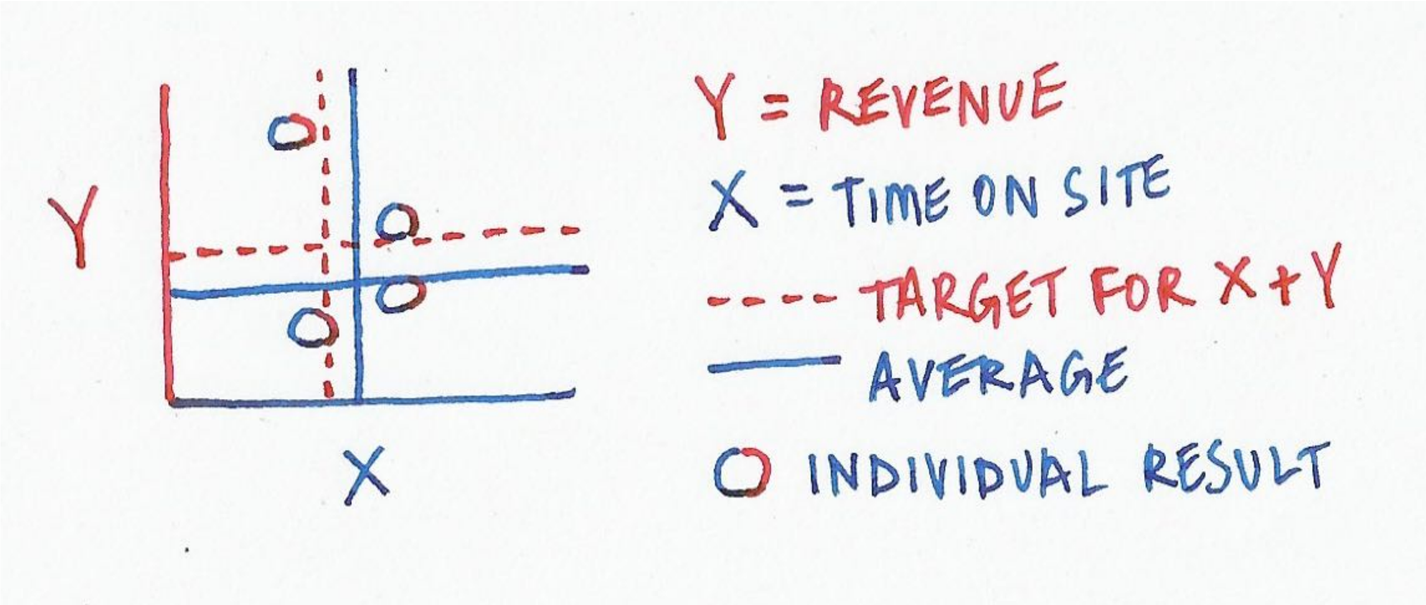
1. **`x`: independent variable** the thing you're testing (potential cause)
2. **`y`: dependent variable** the goal or objective

Ask yourself:

- Does one variable affect another?
- Does one variable affect the objective, or goal?

An example scatterplot

> *Our objective is to maximise $revenue*



The circles are the intersection between `x` and `y`

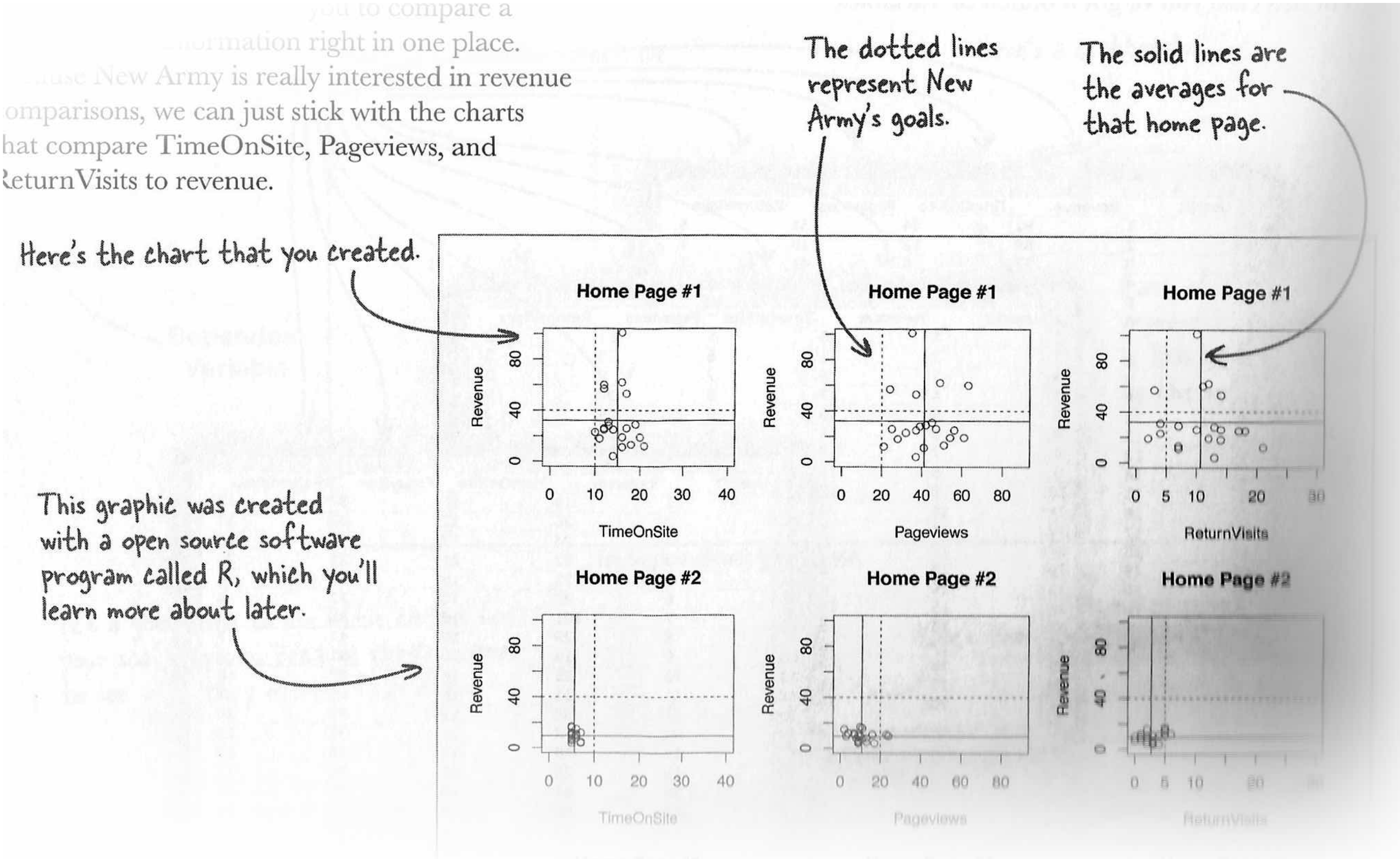
Here we can see our 😀 `customer` spent:

1. `x = 10` minutes on the site
2. `y = 40` dollars worth of stuff

Comparing multiple variables

> *You're just looking for patterns and cause/effects. You don't have to prove it!*

A single visualisation compares the relationship between *two variables*. When looking at variables in data analysis, *more is better* — looking at charts together is a way of testing multivariate variables:



Look around at data visualisations, including the authority on the subject, Edward Tufte. How good a job are they doing?

- **How many variables do they have?**
  - If there are *3 or more* variables, they're more likely to be making intelligent comparisons and summaries.
- **Is it *data art* or *data analysis*?**
  - It isn't data analysis if you can't directly understand the underlying data from the visuals.

05: …

06: …

07: Subjective probabilities

Dealing with probability words

If you have a hunch, a mental model, belief or state, it's easy to say how you feel; but that's often fuzzy. People have differing opinions, or opinions change over time. It's not unusual for people to say things like:

- "It's **highly likely** that cuddly toys will sell next year"
- "I think they **might** prefer to buy ice-cream in summer"
- "Russia will **probably** continue to support oil next quarter"
- "My customers are **more likely than not** going to love this new product line"

These are **probability words**. Probability words don't give a good enough indication how *likely* something is (or isn't) going to happen. They can't be quantified, they're open to disagreement, they're highly subjective.

Ask better questions, get better answers

There is a way we can turn vague, subjective words into **actionable insights**:

1. Look at the topics of conversation. Where are the probability words?
2. Break down the statement(s) into key points
3. Simplify these into questions, that can be asked in a survey
4. Ask participants to answer with a number, or percentage
5. Analyse the results

The answers are still subjective, and are only as good as their input. Misleading or deceptive input is still a risk, but it's a clearer indication of current thinking. In fact, it helps you think.

The results are clearer

probability word	percentage point
certainly	100%
might	50%
likely	70%
definite	99%

Quantifying probability words in this way answers the following questions:

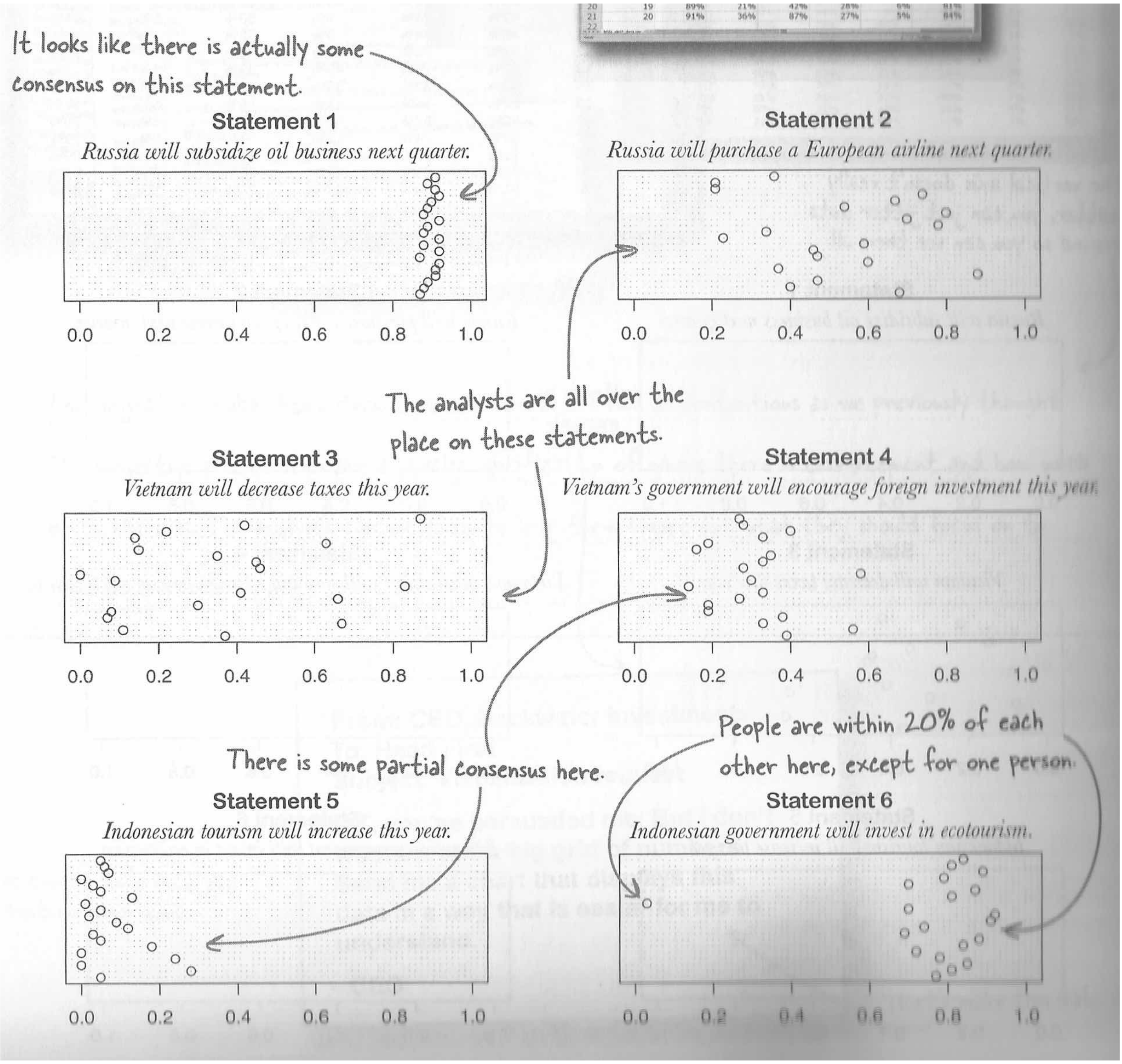
1. Which are hot topics of debate?
2. Which are in agreement (it may be closer than you think!)
3. Which statements need more clarifying or research?

Visualising probability words

> *"What is the probability of X?" — "There's an x% chance of x happening"*

Once you've completed your analysis of the key points, you can easily set up a table for all statements to be analysed. You can then plot these statement results on a graph:

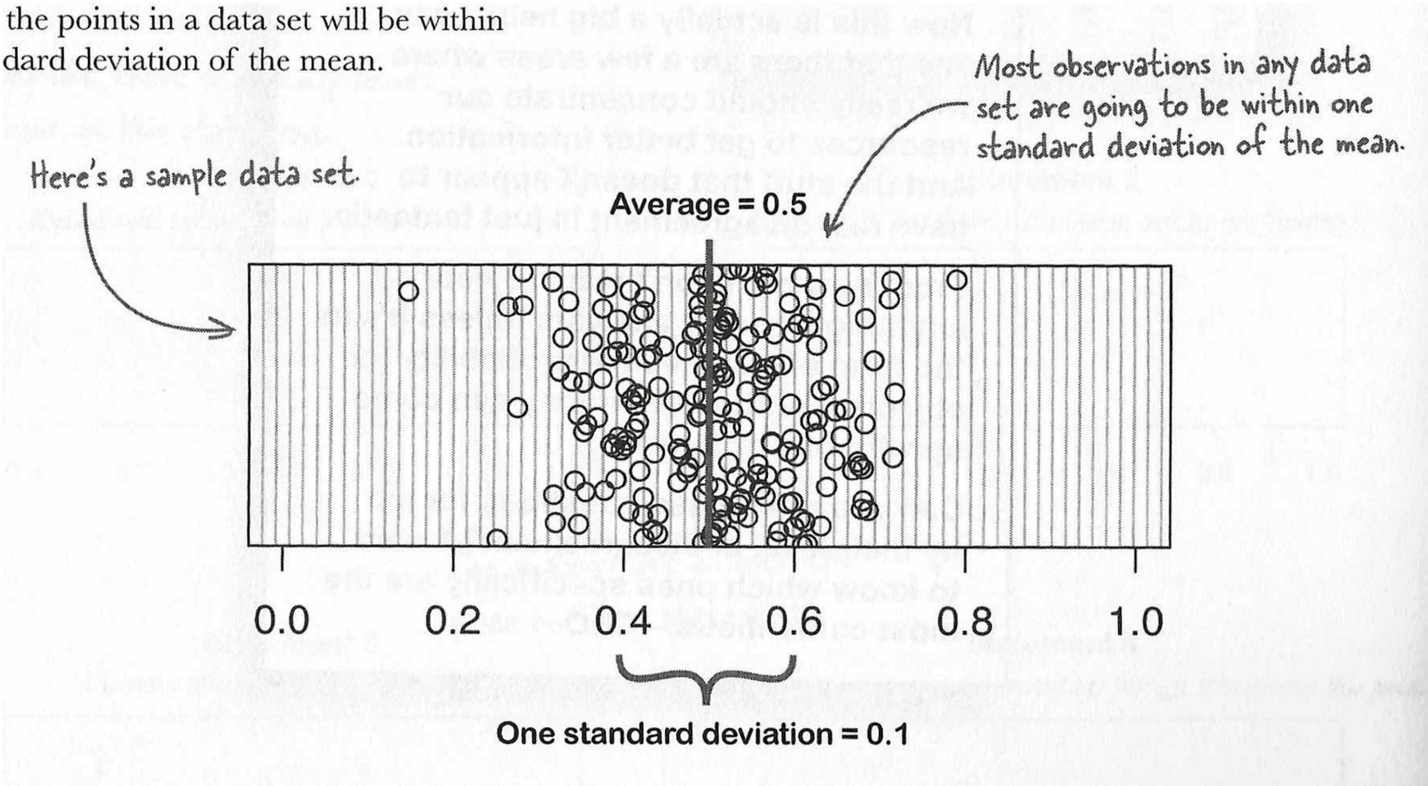
Analyst	Statement #1	Statement #2	…
🤔 Tim	%	%	…
🤨 Bob	%	%	…



Participants are scattered on the vertical, horizontal shows results

Are we in agreement?

Results	🤔 Tim	🤨 Bob	…	`stdev(range)`
Statement #1	%	%	…	%
Statement #2	%	%	…	%



Standard deviation is a great way of answering “how close are the answers to each other?”. It helps us view how far typical points are from the average.

1. **Smaller** standard deviation means **results are closer**
2. **Higher** standard deviation means **they disagree more**

⚠️ It’s still subjective, remember!

This only *describes* the data you’re inputing — it’s still subjective and will (probably) need to be iterated on with new data.

When new information appears

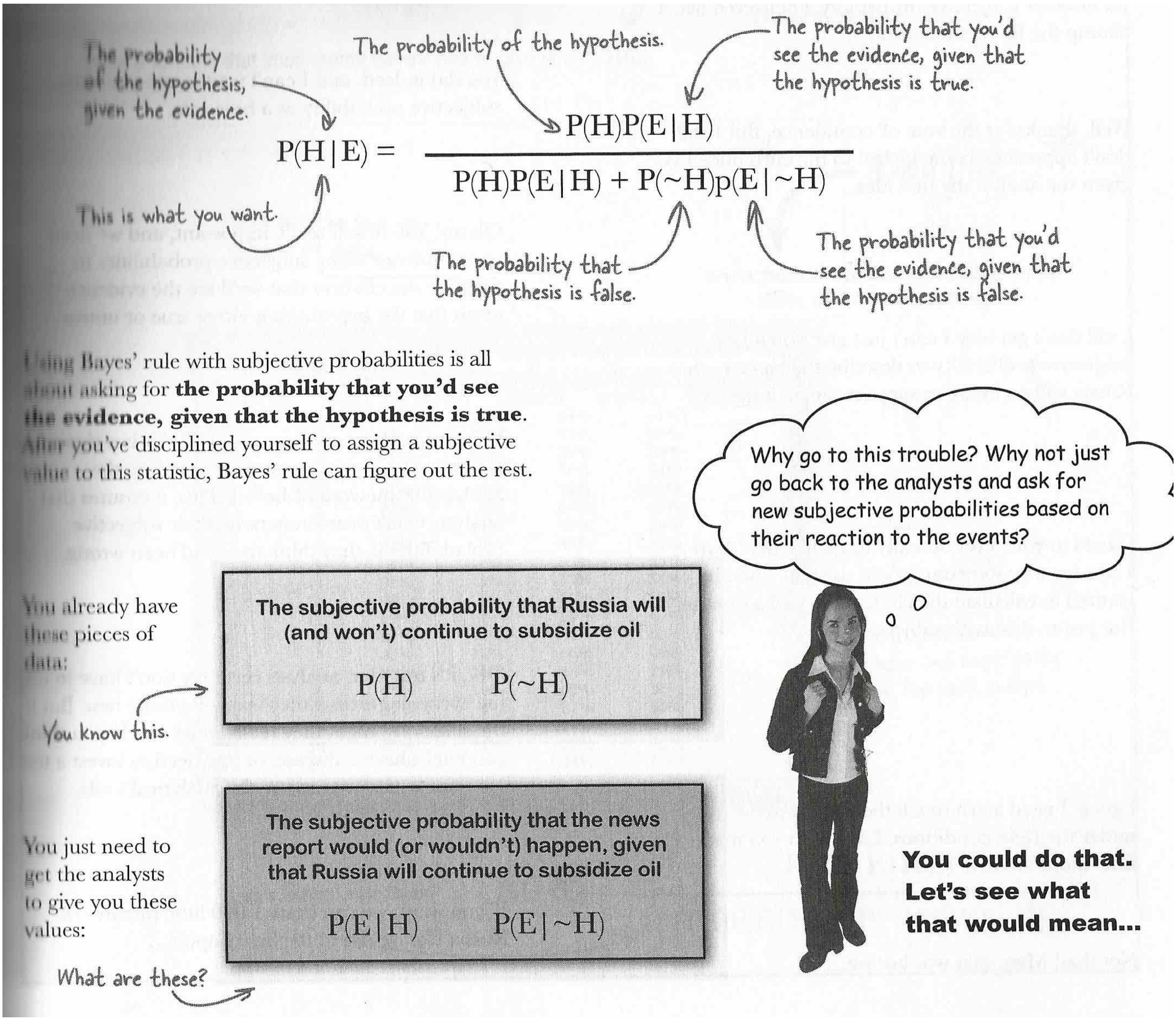
> *Extra! Extra! A new source of information changes the mood …*

When you get new information, or things change, it’s a good idea to reassess people’s assumptions or predictions.

- A new source of information may be reliable, misleading, or false
- It doesn’t mean your analysis is wrong, but it should be adjusted to compensate
    - It *only* describes Truth™ at ⏱️ w/ 💭
- Prepare a new survey, with a subjective probability based on the new intel.

Bayes' rule to the rescue!

If the stakes are high, you don't just want to ask for new subjective probabilities. A more rigorous way of incorporating new info is to use Bayes' rule: this helps avoid people overcompensating for their previous answers.



The probability of the hypothesis, given the evidence.

The probability of the hypothesis.

The probability that you'd see the evidence, given that the hypothesis is true.

$$P(H|E) = \frac{P(H)P(E|H)}{P(H)P(E|H) + P(\sim H)p(E|\sim H)}$$

This is what you want.

The probability that the hypothesis is false.

The probability that you'd see the evidence, given that the hypothesis is false.

Using Bayes' rule with subjective probabilities is all about asking for **the probability that you'd see the evidence, given that the hypothesis is true**. After you've disciplined yourself to assign a subjective value to this statistic, Bayes' rule can figure out the rest.

You already have these pieces of data:

You know this.

The subjective probability that Russia will (and won't) continue to subsidize oil

$P(H)$ $P(\sim H)$

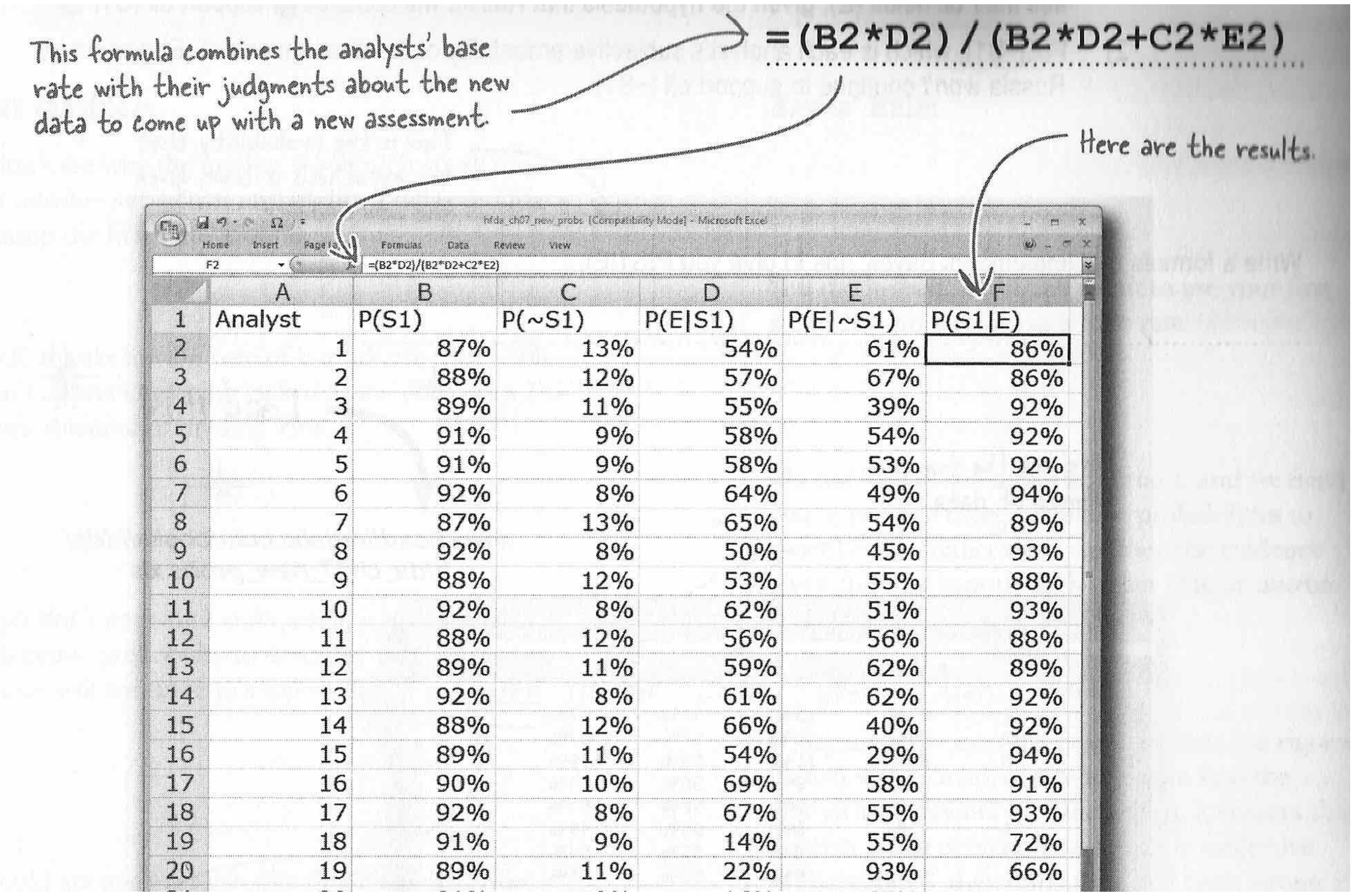
You just need to get the analysts to give you these values:

What are these?

The subjective probability that the news report would (or wouldn't) happen, given that Russia will continue to subsidize oil

$P(E|H)$ $P(E|\sim H)$

Bayes' formula allows you to feed in the new subjective probabilities, using the old ones as a base



This formula combines the analysts' base rate with their judgments about the new data to come up with a new assessment.

=(B2*D2) / (B2*D2+C2*E2)

Here are the results.

F2 =(B2*D2)/(B2*D2+C2*E2)

	A	B	C	D	E	F
1	Analyst	P(S1)	P(~S1)	P(E|S1)	P(E|~S1)	P(S1|E)
2	1	87%	13%	54%	61%	86%
3	2	88%	12%	57%	67%	86%
4	3	89%	11%	55%	39%	92%
5	4	91%	9%	58%	54%	92%
6	5	91%	9%	58%	53%	92%
7	6	92%	8%	64%	49%	94%
8	7	87%	13%	65%	54%	89%
9	8	92%	8%	50%	45%	93%
10	9	88%	12%	53%	55%	88%
11	10	92%	8%	62%	51%	93%
12	11	88%	12%	56%	56%	88%
13	12	89%	11%	59%	62%	89%
14	13	92%	8%	61%	62%	92%
15	14	88%	12%	66%	40%	92%
16	15	89%	11%	54%	29%	94%
17	16	90%	10%	69%	58%	91%
18	17	92%	8%	67%	55%	93%
19	18	91%	9%	14%	55%	72%
20	19	89%	11%	22%	93%	66%

Here you can see the original statement answers; the inverse statement answers; and the new subjective probabilities (given the new evidence)

Comparing results

Once you've run your new results with Bayes, simply create another scatterplot with new results, comparing this with the previous visuals.

1. What does this tell you?
2. Do these new subjective probabilities change your mind?
3. What are the next steps?

08: Heuristics

Analyse like a human

> *The real world is wiggly. It has more variables than you can handle.*

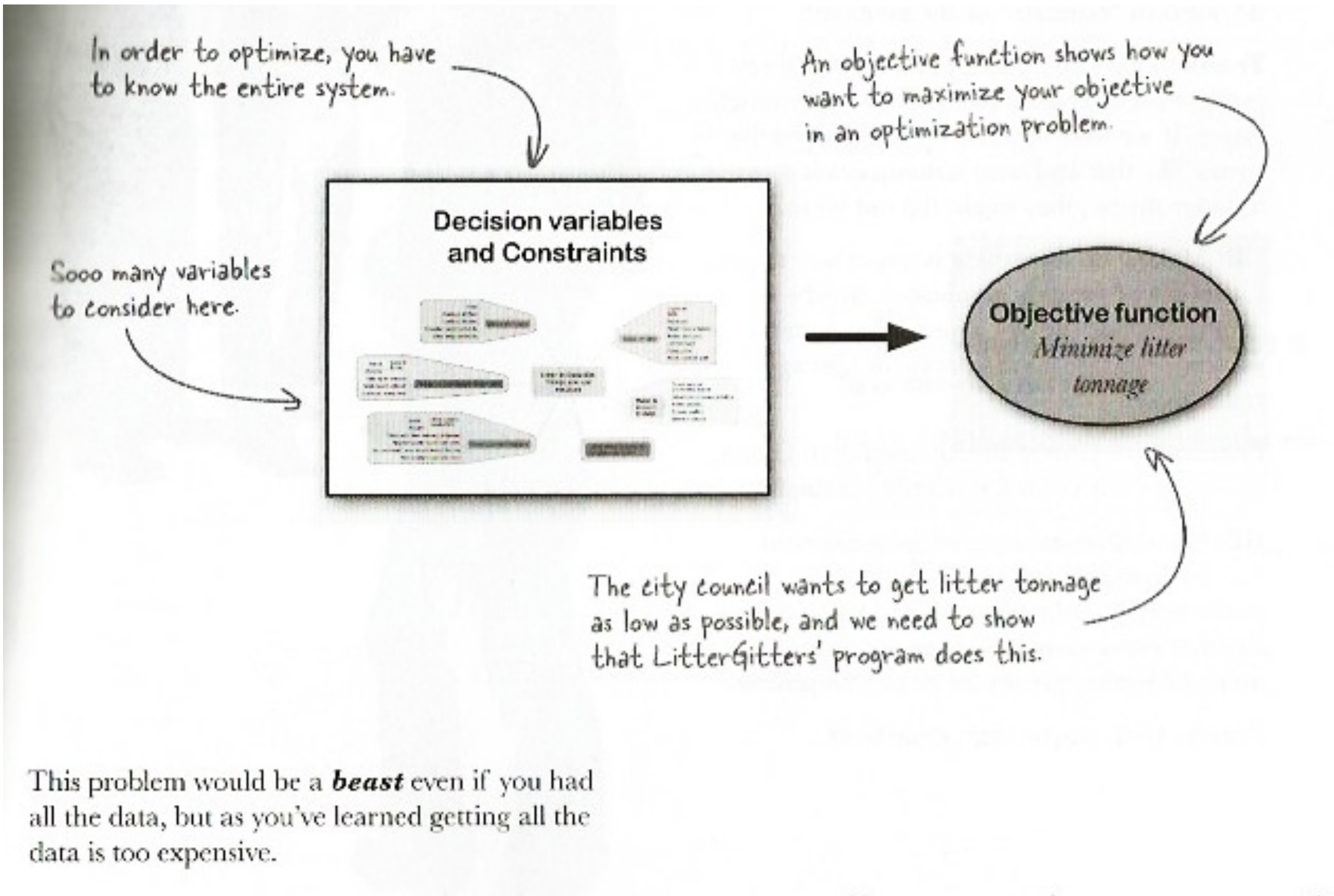
In an ideal world, you'd be able to connect, process and analyse *all* the data you'd need to make a decision. Just tweak a few decision variables and **boom!** you've got the perfect result.

The real world looks like this:

- Many moving parts or variables
- Difficult, expensive, time-consuming to analyse
- Complex mental models
- Too many decision variables

For many tasks, it's **best to simplify**.

It's not an ideal world — keep it simple, stupid!



> *Use the right tool, for the right job*

You can't hold all variables in a complex system. Pick the important ones, understand your solutions are probabilities, not absolutes.

- An optimisation function is useful for ultra-specific problems

- **Use heuristics** in most other situations
- **Pick a sensible objective** and the best variables to achieve it
- **Simplify** wherever possible. It often work

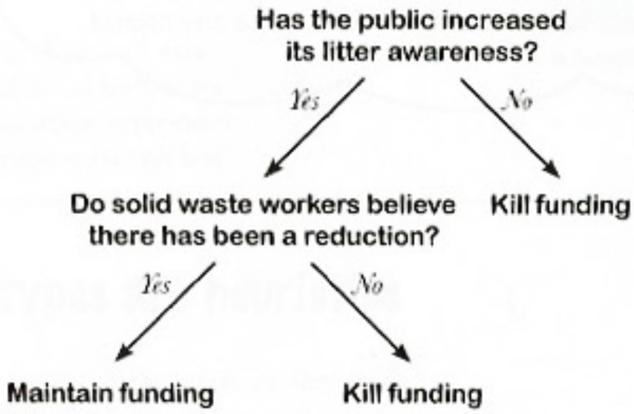
Map out the moving parts of the system. Ask yourself:

1. List and group your system parts
2. What types of data are there? (quantative or qualitative)
3. Which pieces might give clues to achieve my goal?
4. Which are expensive, time consuming, complex, or impossible?
5. Is your goal reasonable? If not, can you describe why?
6. What is realistic, quick, or easy to gather?

Heuristics

explaining what you see to the city council.

Here's how you decided the city council should assess the work of LitterGitters.



Here's our original data describing the attitudes of the general public about litter.

Questions for the general public	Last year	This year
Do you litter in Dataville?	10%	5%
Have you heard of the LitterGitters program?	5%	90%
If you saw someone littering, would you tell them to throw their trash away in a trash can?	2%	25%
Do you think littering is a problem in Dataville?	20%	75%
Has LitterGitters helped you better to understand the importance of preventing litter?	5%	85%
Would you support continued city funding of LitterGitters' educational programs?	20%	81%

Here's some new data describing the sanitation workers' impressions of litter in Dataville since LitterGitters began their program.

Questions for the sanitation workers	This year
Have you noticed a reduction in litter coming into Dataville landfills since LitterGitters began their work?	75%
Are there fewer cigarette butts being collected off the streets since LitterGitters began their work?	90%
Have high-litter areas (downtown, city parks, etc.) seen a reduction in litter since LitterGitters began their work?	30%
Is littering still a significant problem in Dataville?	82%

We can't compare this figure to last year, because we just started collecting data for this report.

These numbers represent the percentage of people who answered "yes."

The brain is fallible. Make it easy on yourself.

1. ~~**Intuition** is seeing one option~~
2. **Heuristics** are seeing a few options
3. ~~**Optimisation** is seeing all options~~

Most decision making uses heuristics; we can use a **fast and frugal tree** to break a problem into manageable pieces, figure out easier ways to make a decision:

- Far more efficient
- Cost less in time, money or mental effort

- Simple, and often work well

Who are you trying to convince?

> *The council don't care about your public opinion survey. They want hard data, not fuzzy feelings.*

Remember, your client also has their own mental model:

- How does your customer think?
- What drives their decisions?
- Are they what you think?
- Are they realistic?

It's your job to convince them

> *The goal they're asking for isn't feasible. It's too hard.*

Understand your customer's decision making process — use this as a basis for your own analysis, helping them decide. If the goal is impossible to analyse, could you do it by proxy?

- Agree on a clearly defined outcome
- What facts or figures would you need?
- Do they prove your point?
- … or disprove it? (be honest)

Appendix

Further reading

- **Five hows** and **Five whys** (chapter 01)
- How do you **weight** opinions? variables? results?
- What other examples of **Bayes law** can you find?