UVA DS2003 Communicating with Data

Fall 2026

Infographics

Telling a story with a montage of figures

Sarah Groves

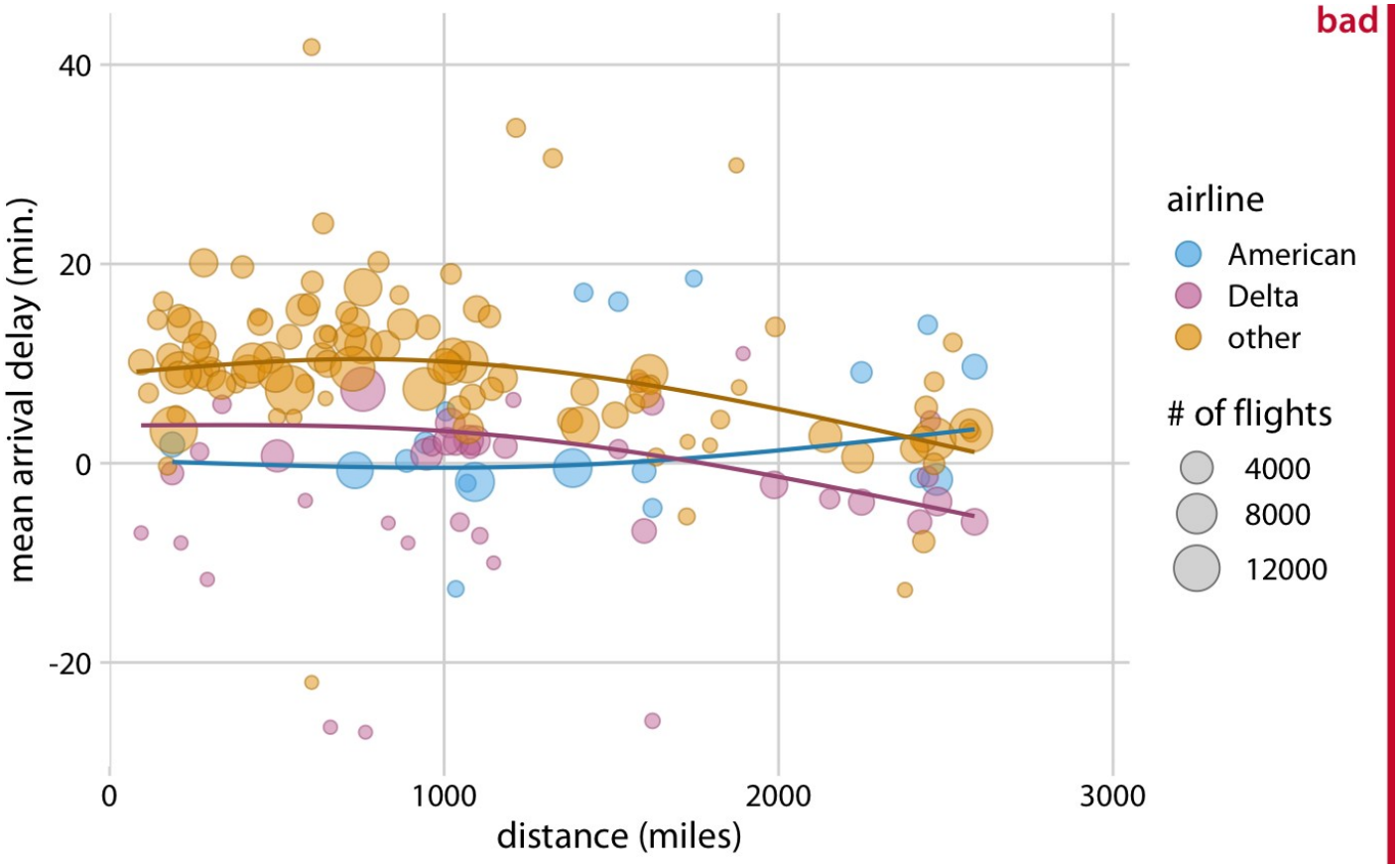
based on Claus O. Wilke, Fundamentals of Data Visualization, ch. 29: “Telling a story and making a point”

Part 4

Rules for figures in a story

Five habits that keep the audience with you

1. Make a figure for the generals



TOO MUCH

Wilke's Army program managers: a general should see the point of each figure at once.

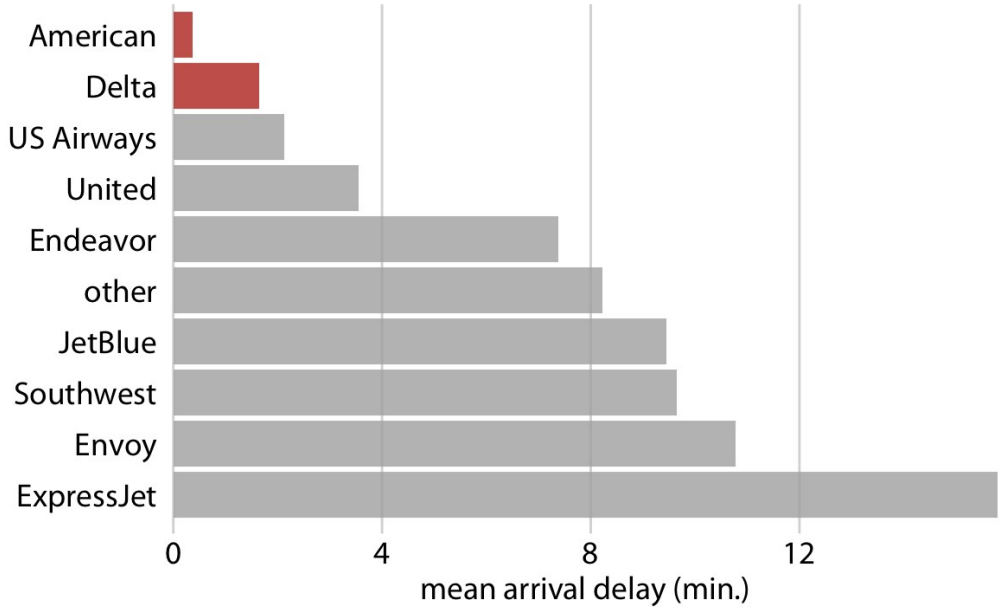
Generals aren't slow. They're busy. So are reviewers, editors, and your boss.

Never assume your audience can rapidly process complex displays.

Figure 29.3: Claus O. Wilke, Fundamentals of Data Visualization (O'Reilly, 2019), ch. 29. clauswilke.com/dataviz Labeled "bad" in the book.

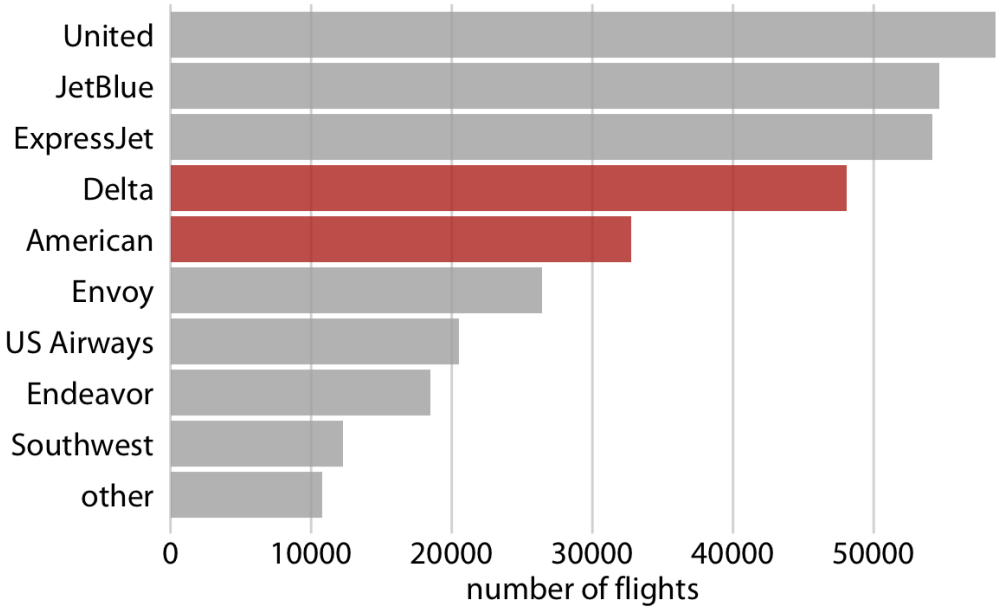
Show the point then answer the next question

THE POINT



Delta and American have the lowest delays.

THE FOLLOW-UP



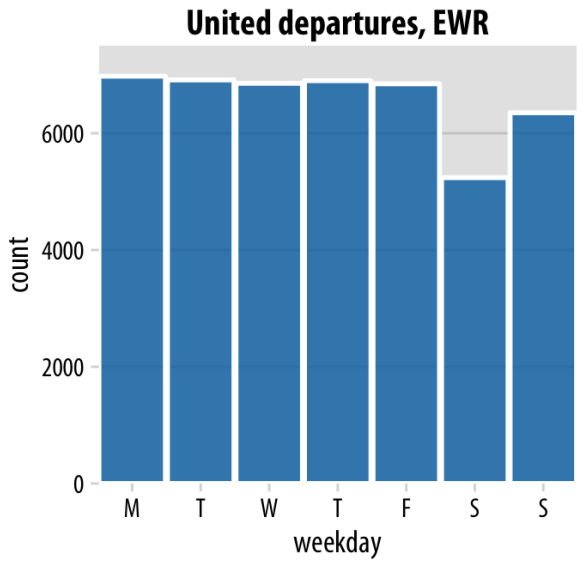
…and they are big NYC carriers, not tiny ones.

Both drop distance. That's fine: don't show dimensions that are tangential to your story.

Figure 29.4 and 29.5: Claus O. Wilke, Fundamentals of Data Visualization (O'Reilly, 2019), ch. 29. clauswilke.com/dataviz

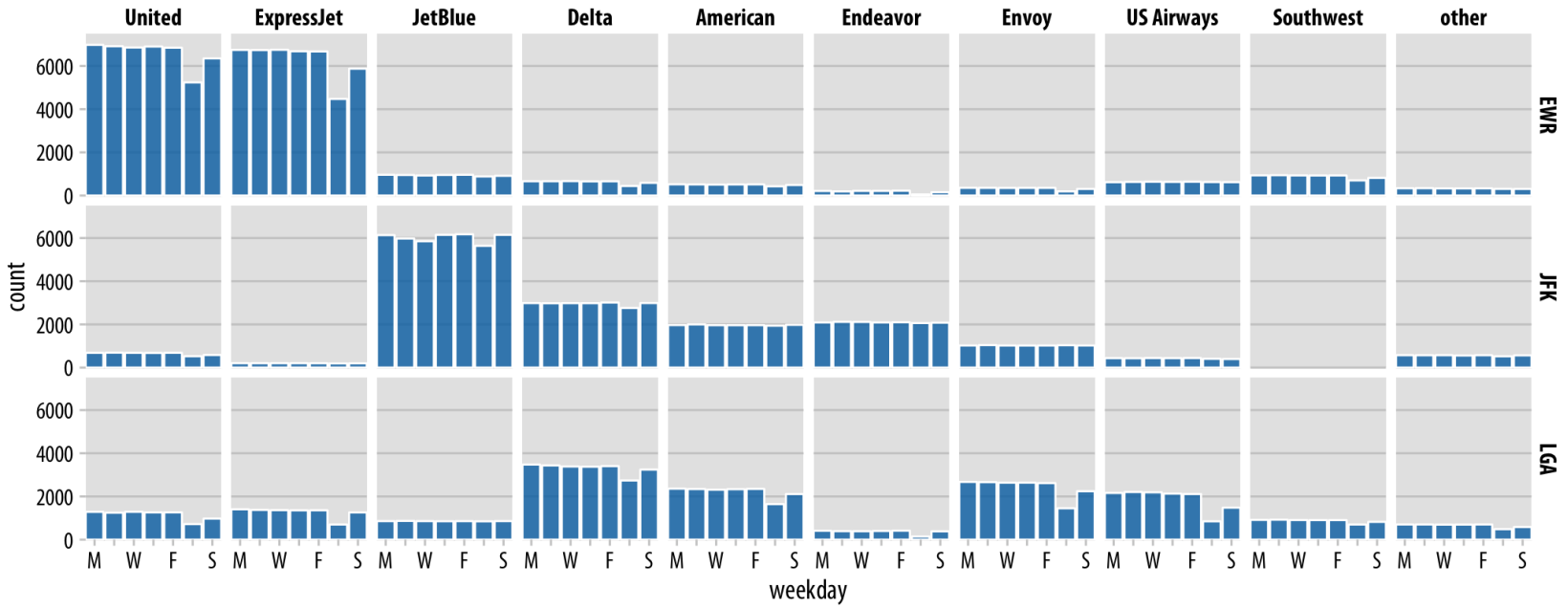
2. Build up to complex figures

FIRST



United departures from Newark, by weekday

THEN



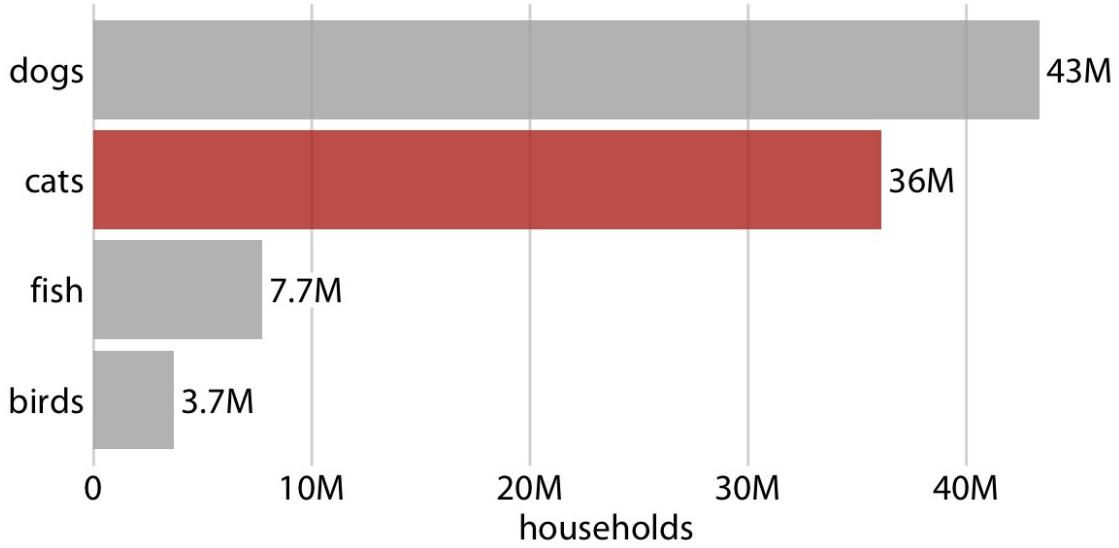
Every airline × airport, same panel design

Never jump straight to a complex figure. Show one panel of a small-multiples grid first; the full grid is then easy to read.

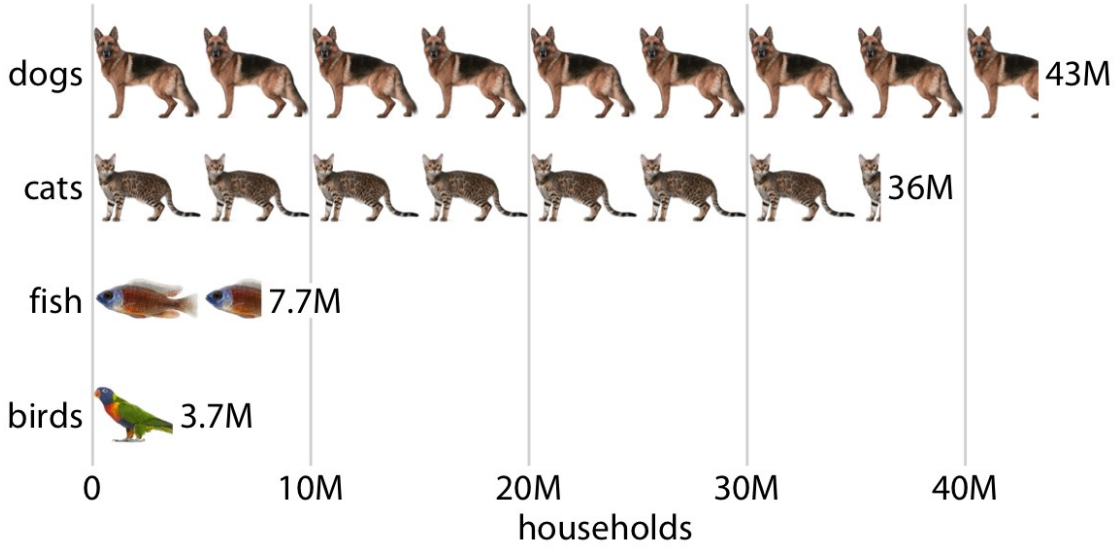
Figure 29.6 and 29.7: Claus O. Wilke, Fundamentals of Data Visualization (O'Reilly, 2019), ch. 29. clauswilke.com/dataviz

3. Make figures clear and memorable

CLEAR BUT GENERIC



CLEAR AND MEMORABLE

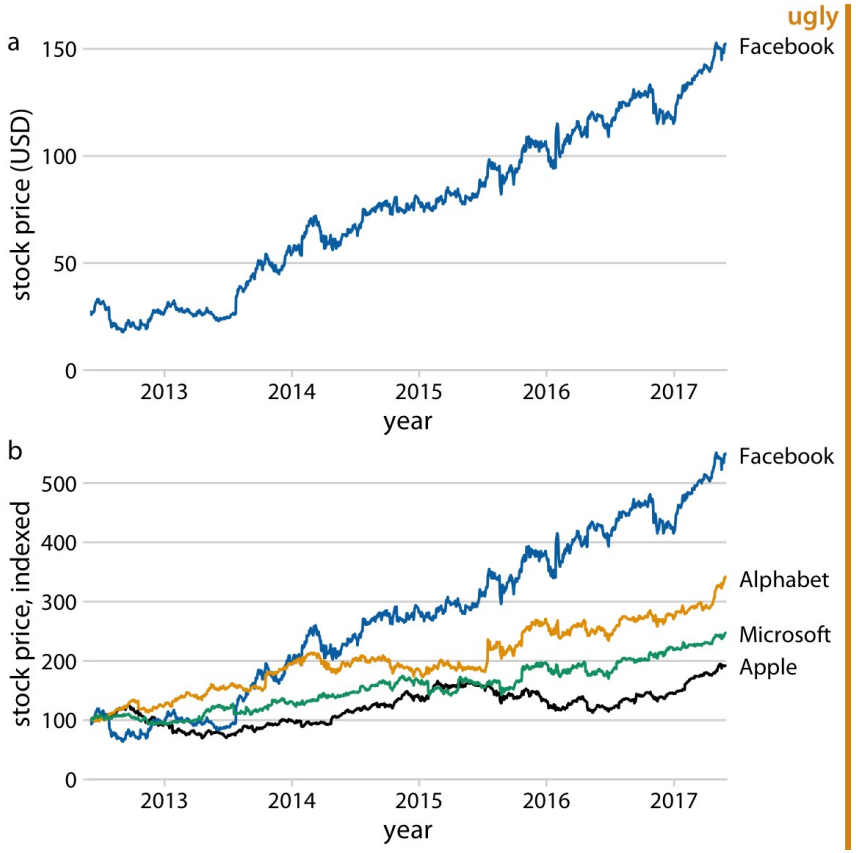


- An isotype chart is still a bar chart: each whole animal = 5 million households.
- Images should be the data, not decorate the axes. Decorative images distract (Haroz et al., 2015).

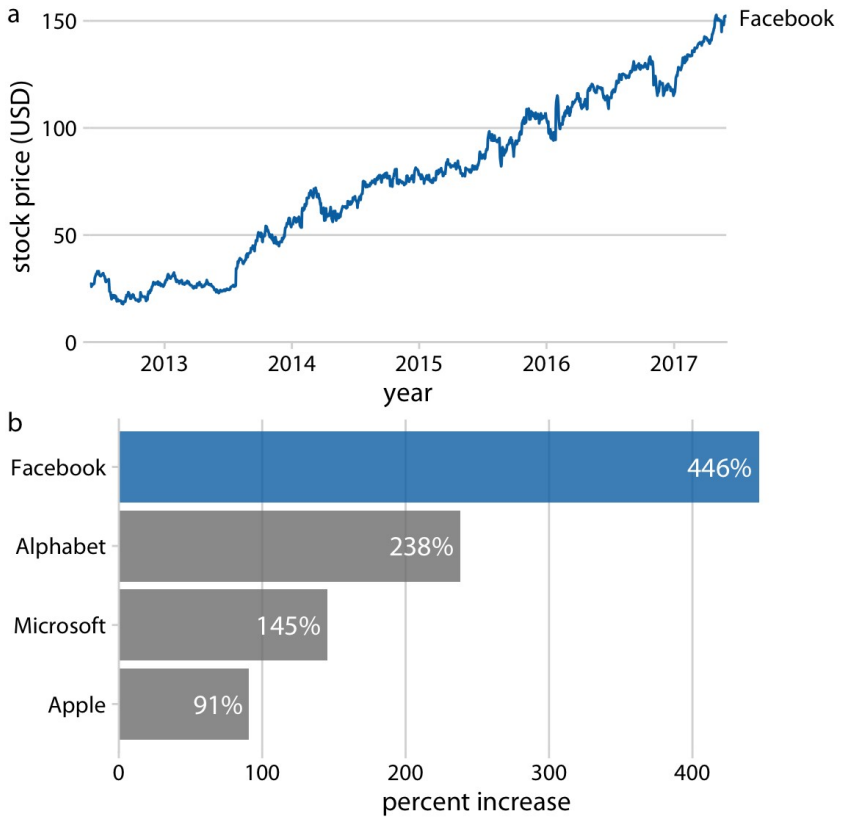
Figure 29.8 and 29.9: Claus O. Wilke, Fundamentals of Data Visualization (O'Reilly, 2019), ch. 29. clauswilke.com/dataviz Data: AVMA, 2012.

4. Be consistent, but don’t repeat yourself

REPETITIVE



DISTINCT



Same visual language (colors, fonts), different chart types for different analyses.

Panel (b) only needs to say “Facebook grew more”: a bar of % increase does that.

Raw data first, derived quantities later.

Figure 29.11 and 29.12: Claus O. Wilke, Fundamentals of Data Visualization (O'Reilly, 2019), ch. 29. clauswilke.com/dataviz Data: Yahoo Finance.

5. Check the uncertainty before you claim it

- Before you report a result, ask how sure you are: sample size, variation, missing data, confidence intervals.
- Your infographic doesn't need error bars or CIs, but your main conclusion has to survive them.
- If a difference is inside the noise, it isn't your headline.
- Your plots must not mislead. (More on deceptive plots later in the semester.)

Holds up

Claim: Friday the 13th has 5.1% fewer births than other Fridays.

Check: Single Fridays swing about ±7%, but 23 of the 25 Friday the 13ths came in below average.

Doesn't hold up

Claim: Library Science is the worst-paid major ($22,000).

Check: That median comes from 2 survey respondents.

Most importantly…

Simplify,
simplify,
simplify!

Simplicity is about subtracting the obvious and adding the meaningful.

Think | Pair | Share

Think Find one infographic around SDS

Pair What's the main story? What story format does it use?
Does it follow or break the rules we just learned?

Share Take a picture of the infographic. Report back to the class.

Add your picture and short summary to the Discussion board!!

Activity

Sketch a data story

Paper and pens only. No code.

Sketches are…

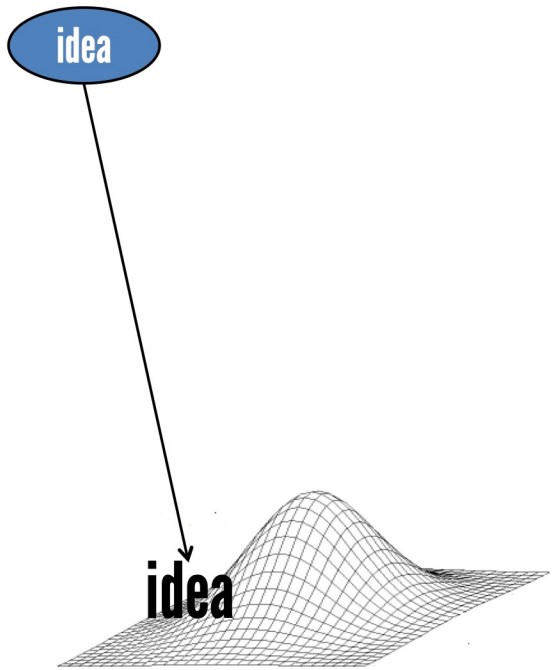
- Quick, freehand drawings
- Can include labels or captions
- Don’t need to be pretty
- **Goal**:
  - For communication
  - For brainstorming

GOAL:
try to communicate ideas with as few lines (as little “ink”) as possible!

Why Sketch?

getting the design right

- generate an idea

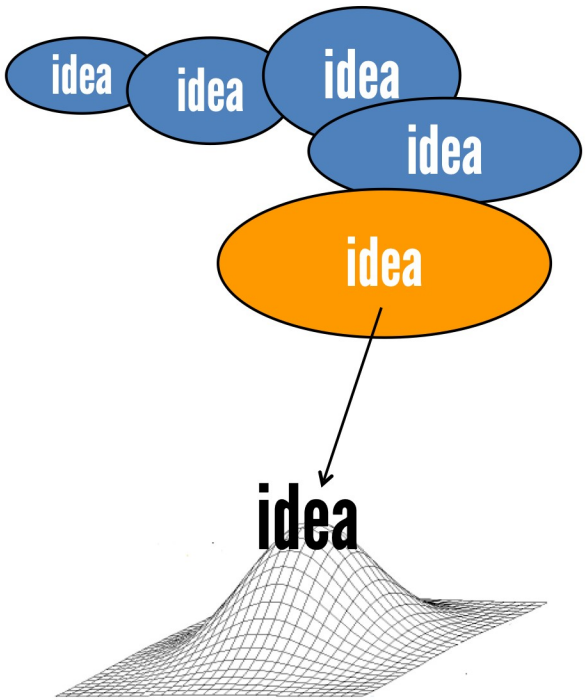


Graphs on the next few slides were modified from images found in the Sept. 2009 Wikipedia entry: *hill climbing*

Why Sketch?

getting the design right

- generate an idea
- iterate and develop it

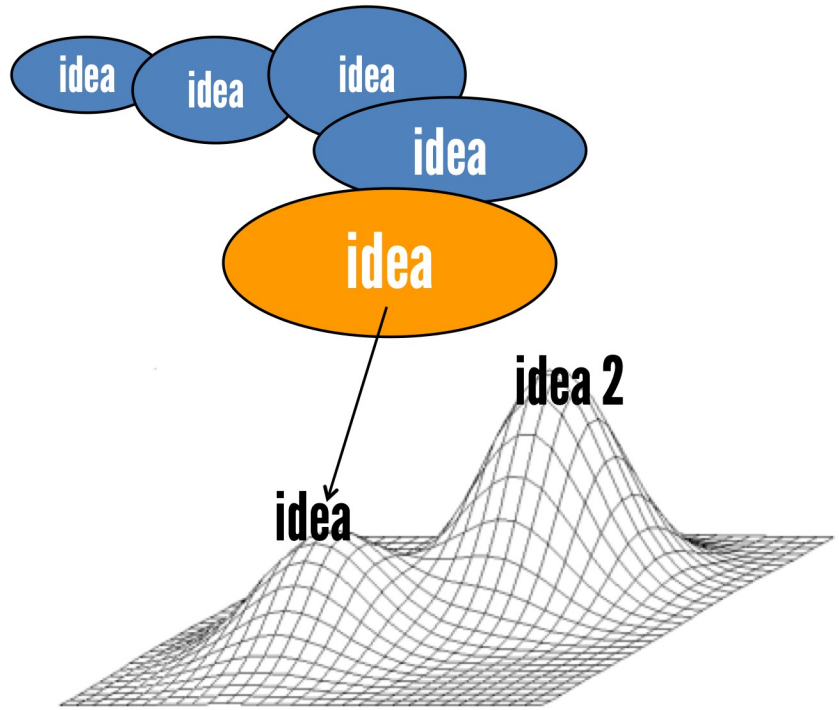


but is it the best idea?

Why Sketch?

getting the design right

- generate an idea
- iterate and develop it



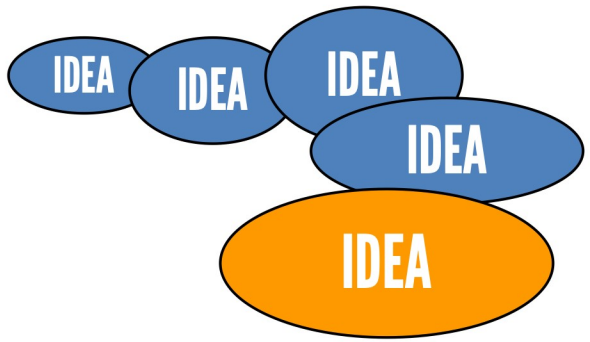
The problem

- other better solutions may be available in different ideas
- local *vs.* global maxima (local hill climbing)
- often results from fixating on a single idea

Why Sketch?

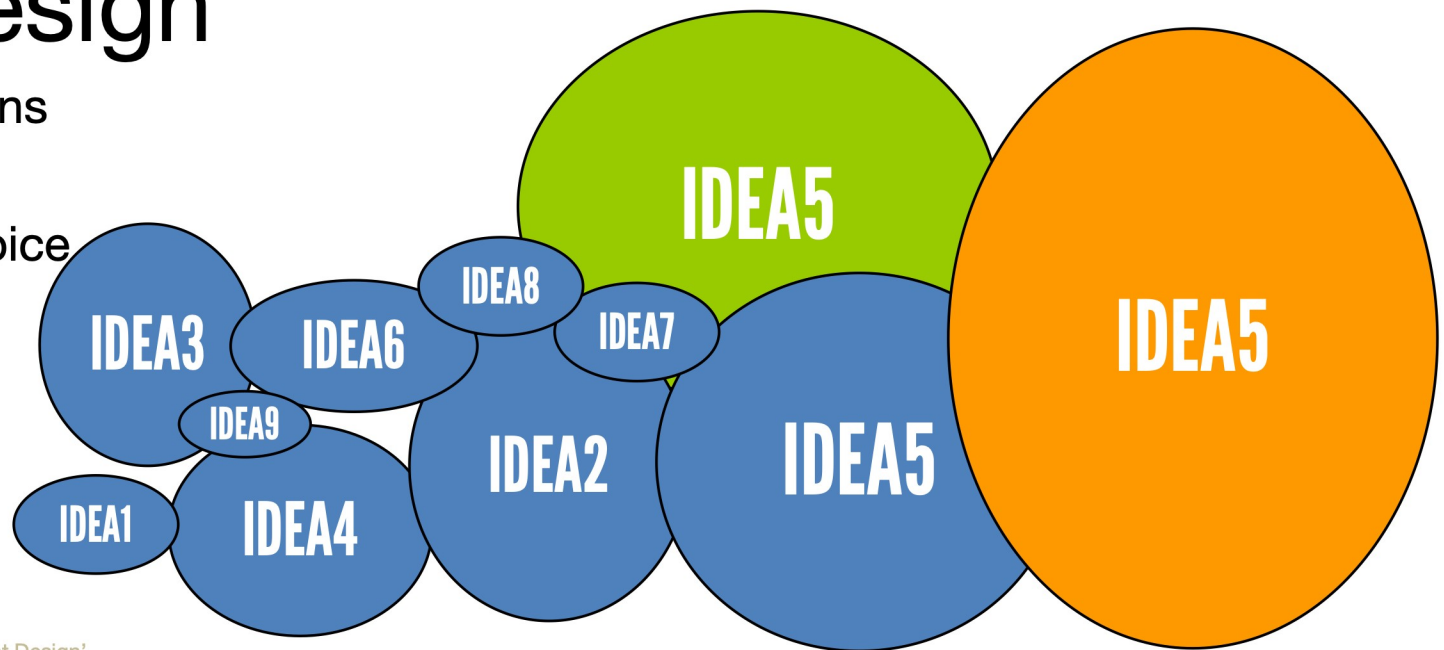
getting the design right

- generate an idea
- iterate and develop it



getting the right design

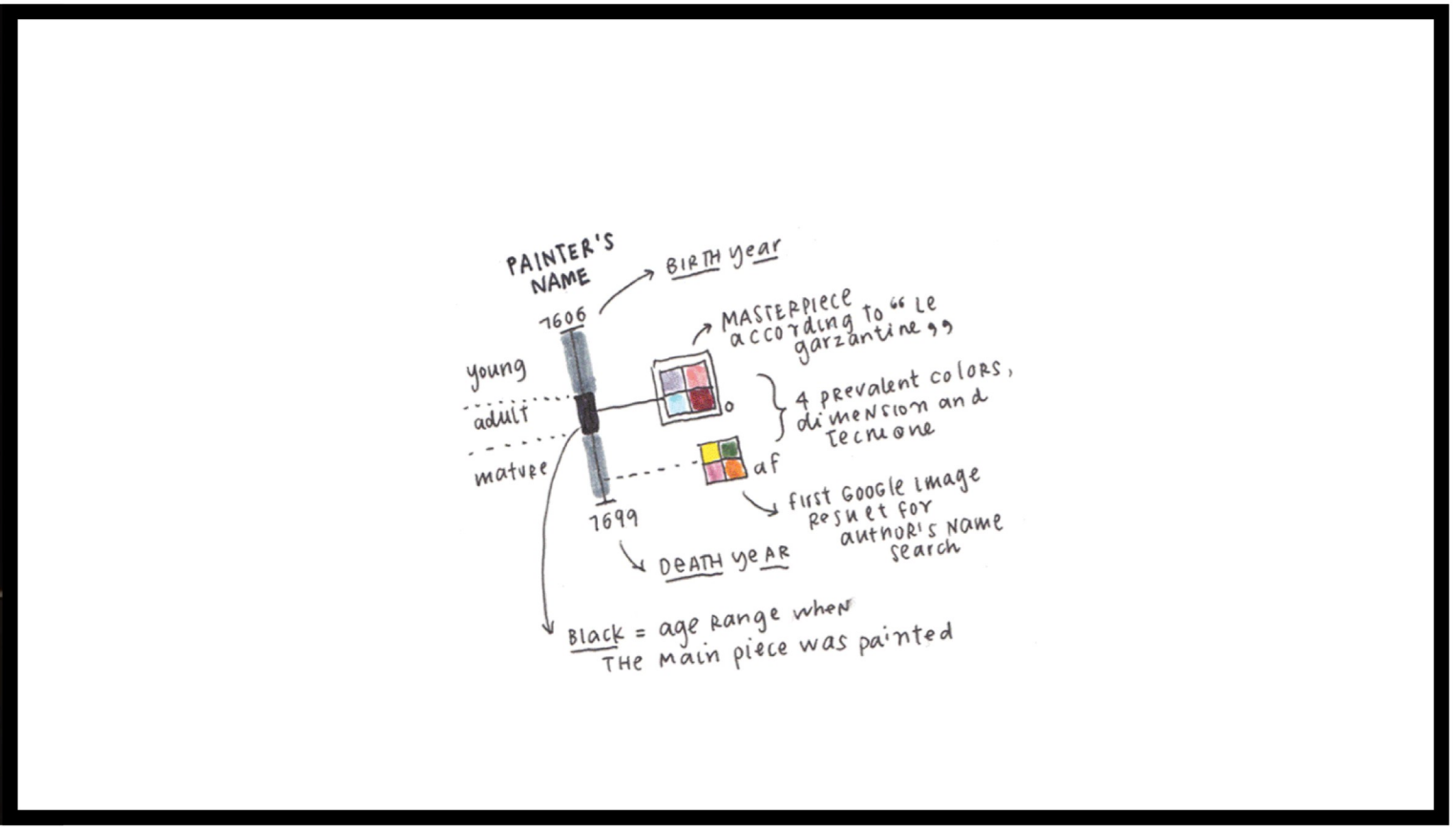
- generate many ideas and variations
- reflect and choose
- *then* iterate and develop your choice



Bill Buxton coined the expression 'Getting the Design Right *vs.* Getting the Right Design'

https://giorgialupi.wordpress.com/2014/07/01/eyeo-2014-the-shapes-of-my-thoughts/





Giorgia Lupi - The Shape of My Thoughts

Watch the full talk: https://vimeo.com/channels/eyeo2014/109954123


BIG TITLE
GUEST NIGHTS BOOKED
LISTINGS ON AIRBNB
3-4 HOST INCOME FACTOIDS GO HERE
HOSTS


GUEST NIGHTS BOOKED
5 MILLION NIGHTS BOOKED
500%
2011
19,000+
192
UK 748%
FRANCE 425%
ITALY 948%
SPAIN 718%
JUST LAST YEAR...
RIGHT NOW...
AIRBNB TRAVELER NETWORK
—NETWORK EFFECTS—
2012
AND THE CYCLE BEGINS AGAIN


Longest consecutive reservation
91 nights
Properties Listed
Positive reviews left
Nights Stayed
THEN
2008
2009
2010
2011
Notable event
Notable event
Quote
Quote
Quote
Building Culture


A SNAPSHOT from FEBRUARY 2012
airbnb
GUEST NIGHTS BOOKED
5M NIGHTS BOOKED!
2011
25%
33%
42%
U.S. Guest
Non-U.S. Guest
LISTINGS ON AIRBNB

AIRBNB

BRINGING THE WORLD TOGETHER

1

10M
8M
6M
4M
2M

NIGHTS BOOKED JUNE

5M JUST 5 MONTHS AGO!

JUNE 08 JUNE 09 JUNE 10 JUNE 11 JUNE 12

10 MILLION GUEST NIGHTS BOOKED

2 A NIGHT ON AIRBNB IS BOOKED...

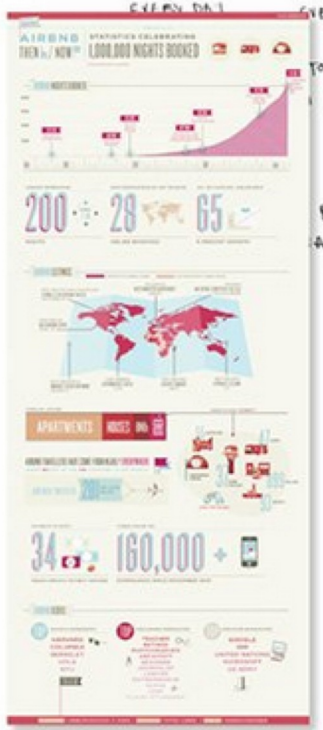
2008: EVERY DAY
2009: EVERY FIVE MINUTES
2010: EVERY MINUTE
2011: EVERY 10 SECONDS
2012: EVERY 2 SECONDS

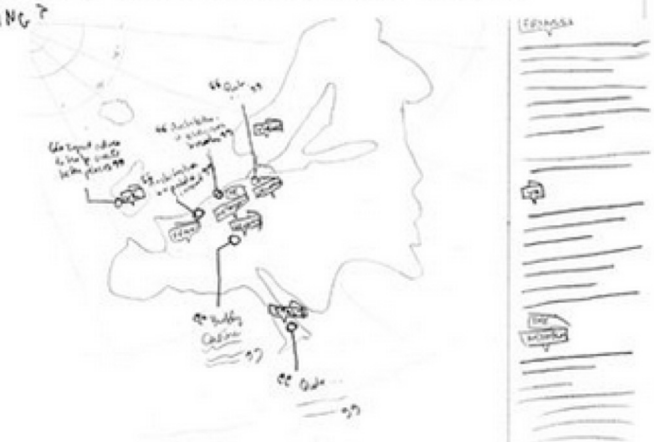
TOOK YOU TO READ THIS, SOMEONE BOOKED A NIGHT'S AIRBNB. IN FACT, IT WAS THIS GUY:

HE'S GOING TO _____ (CITY BASED ON COUNTRY)

READ THIS, 38,000 PEOPLE ARE STAYING ON AIRBNB.

AYING?





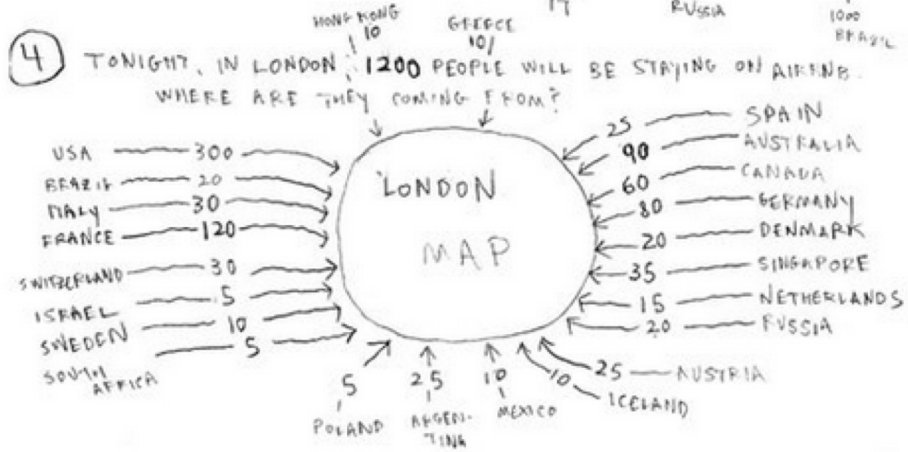
WHERE ARE THEY FROM?

38,000 STAYING ON AIRBNB RIGHT NOW

15,000 USA
2000 FRANCE
1000 ES
1000 IT
UK 3000
GERM 2500
1000 RUSSIA
CANADA 2500
1000 BRAZIL
2500 AUSTRALIA

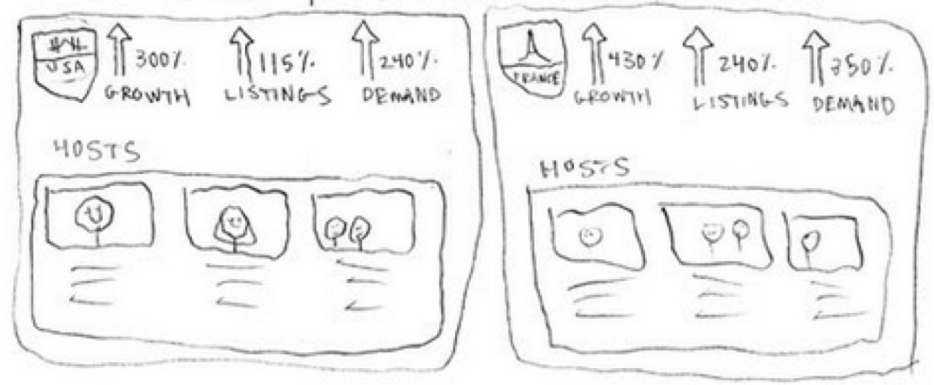
4 TONIGHT, IN LONDON, 1200 PEOPLE WILL BE STAYING ON AIRBNB.

WHERE ARE THEY COMING FROM?

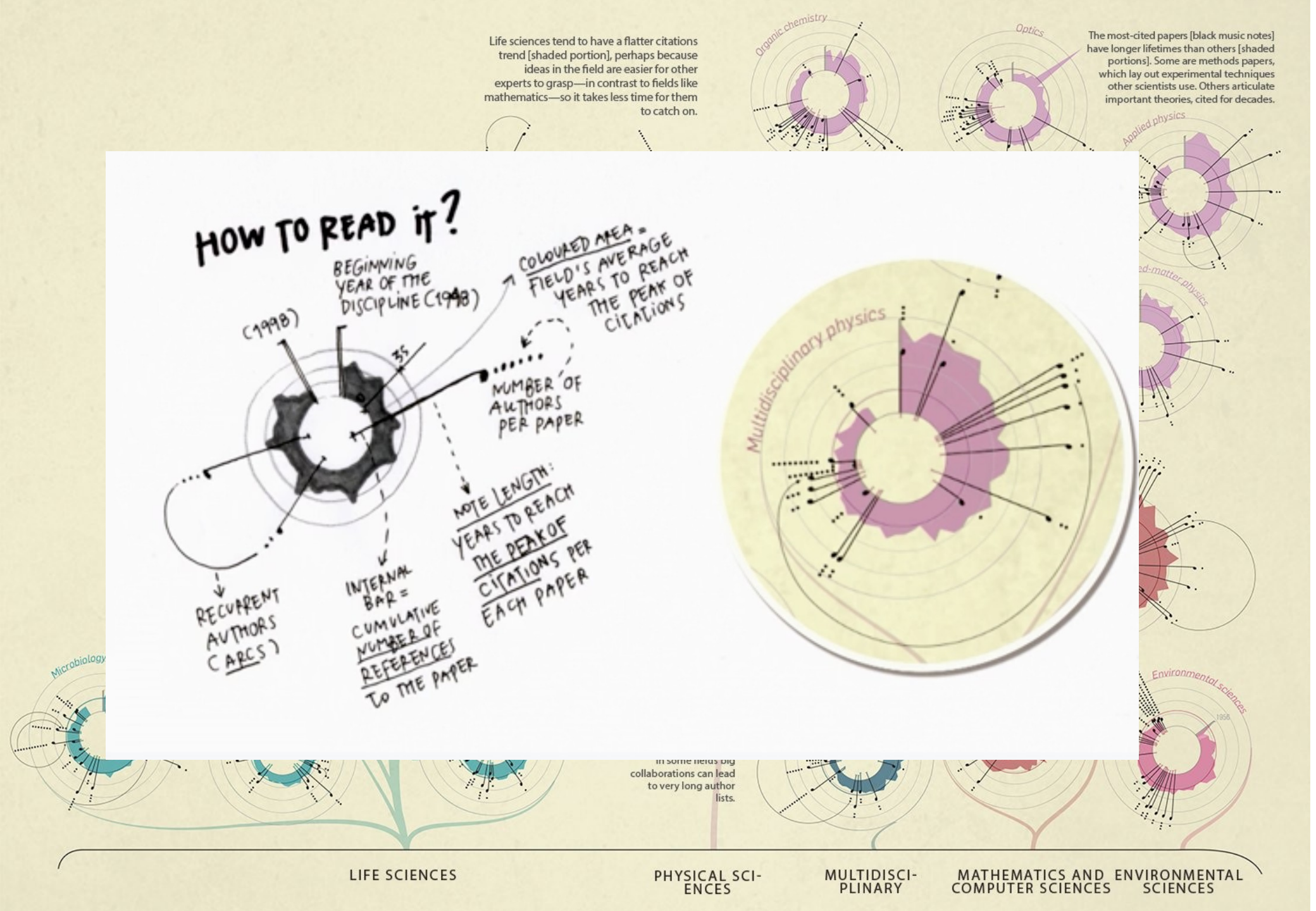


5 INTERNATIONAL GROWTH -

IN THE LAST YEAR, THE AIRBNB COMMUNITY HAS GROWN...



OTHER COUNTRIES



https://www.popsci.com/article/science/life-

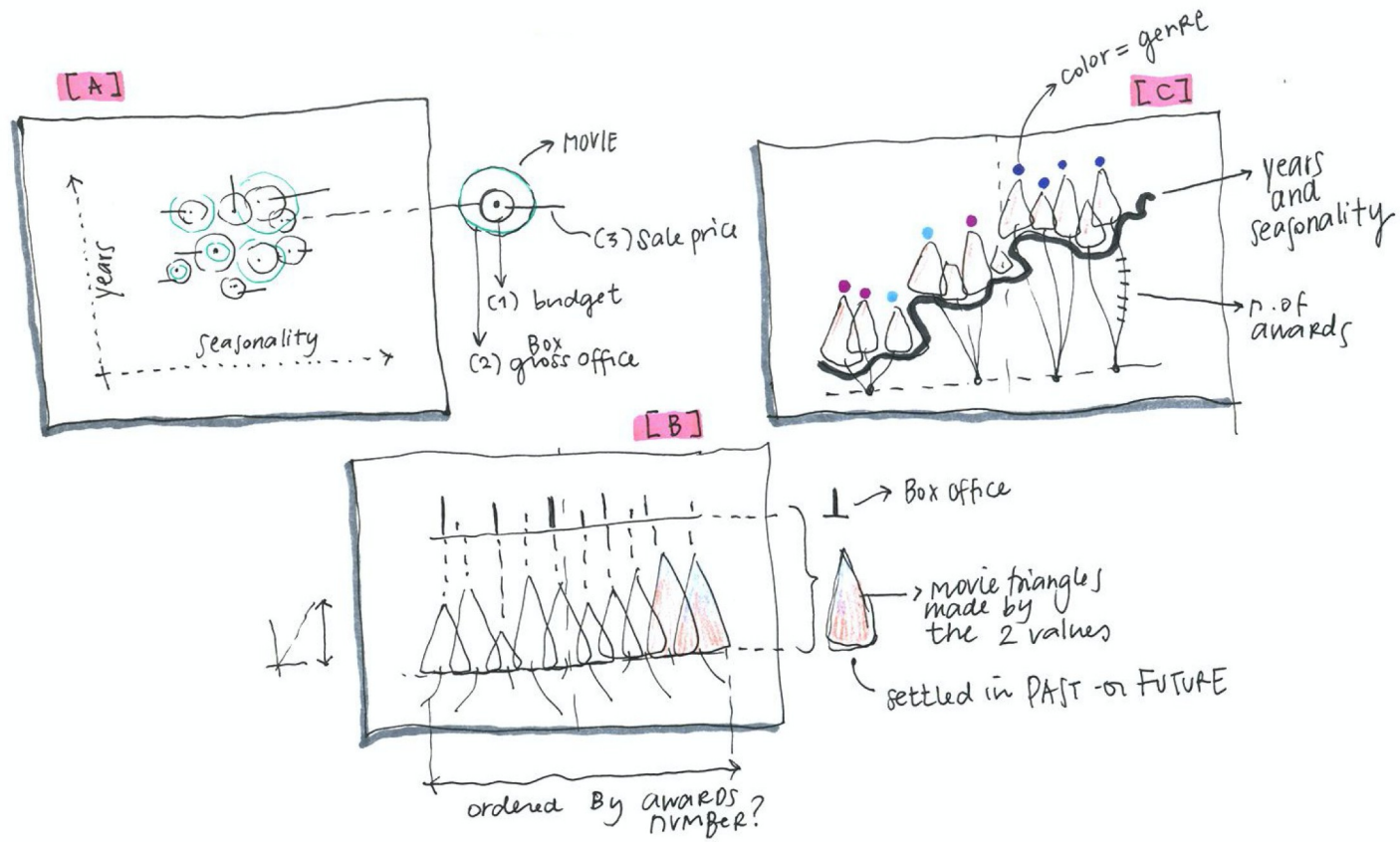
Sketching with Data Opens the Mind's Eye by Giorgia Lupi

"How about the real data? When do the actual data come to the table in the process of sketching a data visualization?"

- Lupi's 3 Phases
  - Phase 1: Overview
  - Phase 2: Data Points
  - Phase 3: Prototype

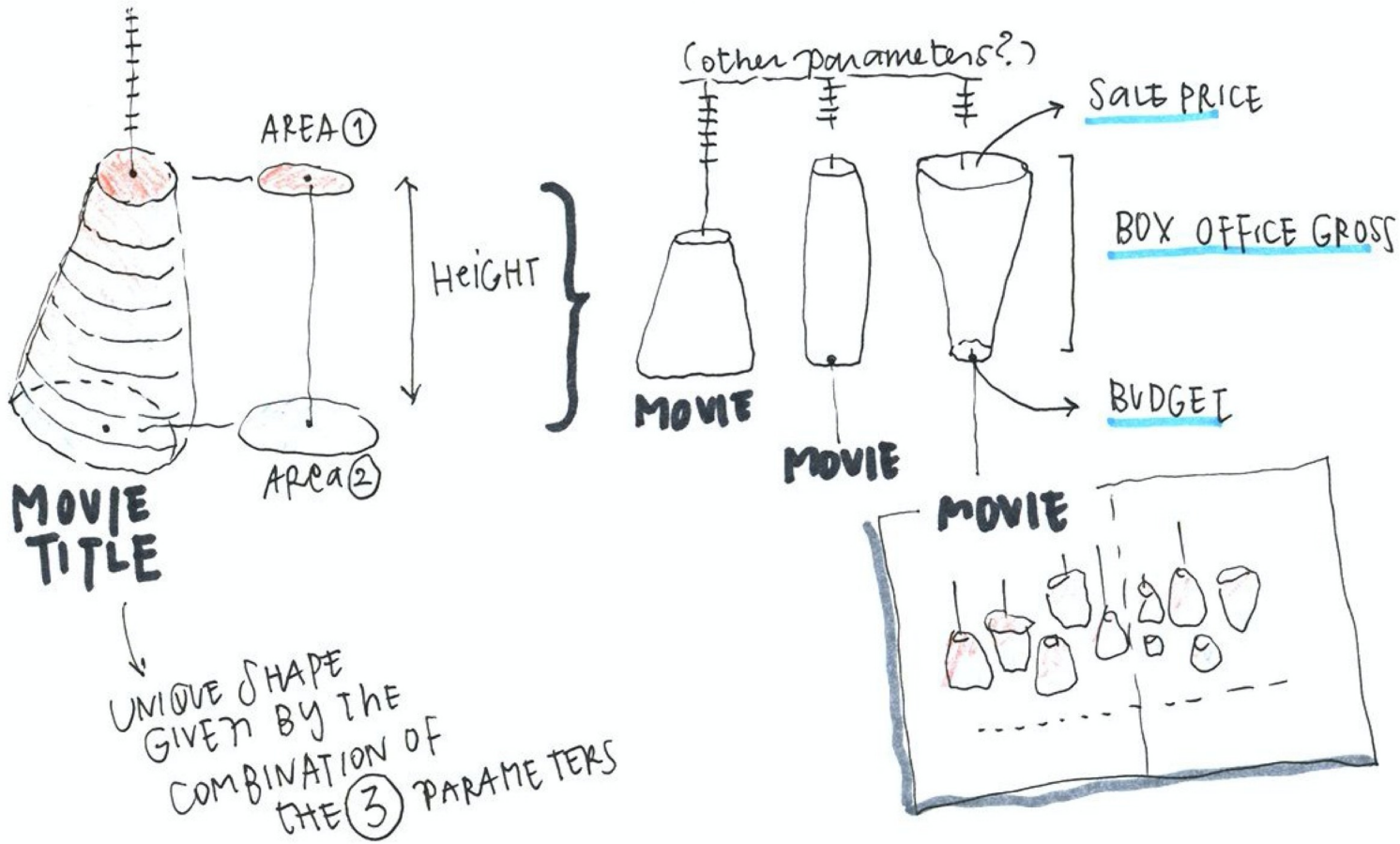
Sketching with Data Opens the Mind's Eye by Giorgia Lupi

- Phase 1: Focus on Overview (overall architecture of the information)



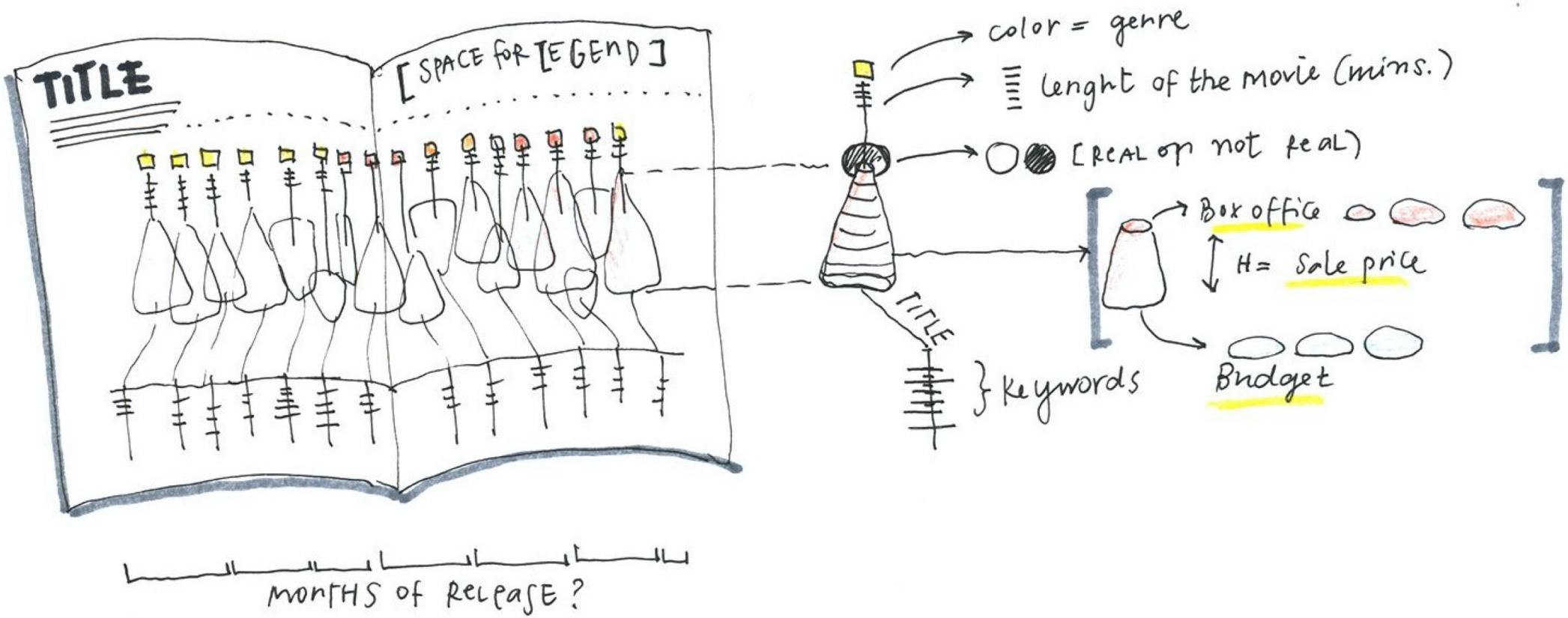
Sketching with Data Opens the Mind's Eye by Giorgia Lupi

- Phase 2: Focus on Data Points (shapes, colors, features to represent data)



Sketching with Data Opens the Mind's Eye by Giorgia Lupi

- Phase 3: Prototype Visualizations (Paper or Digital) to facilitate communication to others



Sketching Workshop

BUT... *I CAN'T DRAW*

Some principles for sketching

- use as few lines as you can
- communicate the essence of the idea
- details only if they are important
- choose the detail you put in deliberately
- one sketch per page!!!!!

SKETCH THE RELATIONSHIP BETWEEN TWO NUMBERS

75 37

(there are at least 45 different ways)

between two quantities

Notation variants

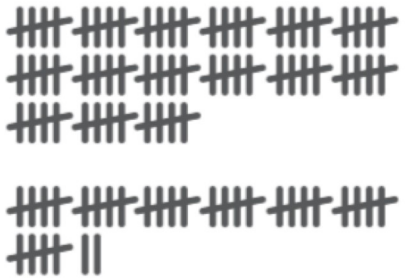
75, 37

a



b





c

between two quantities

Squares



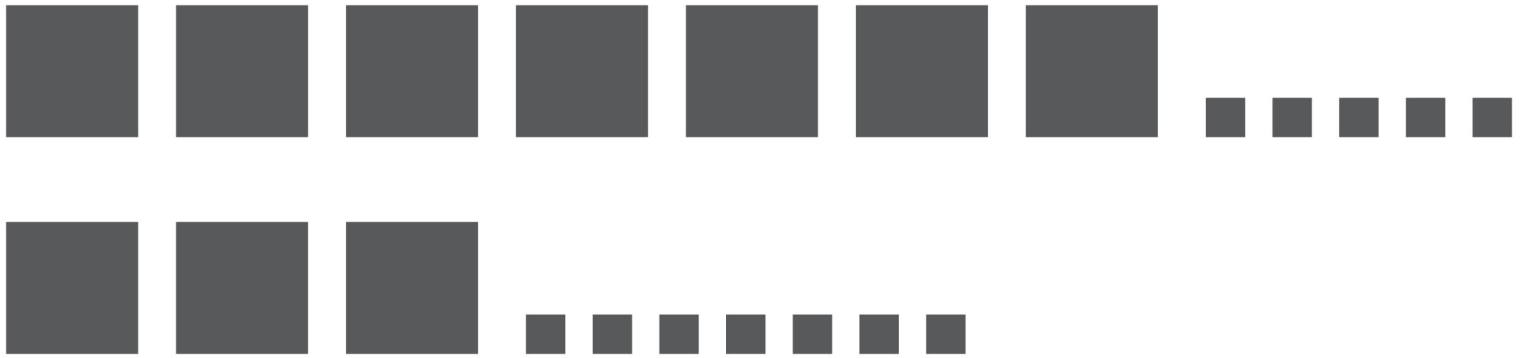
between two quantities

Repeated Icons



between two quantities

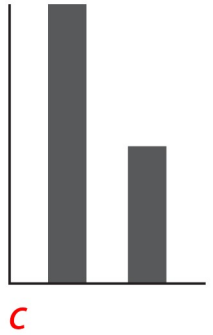
Hundreds, tens, units, decimals … represented by squares



between two quantities

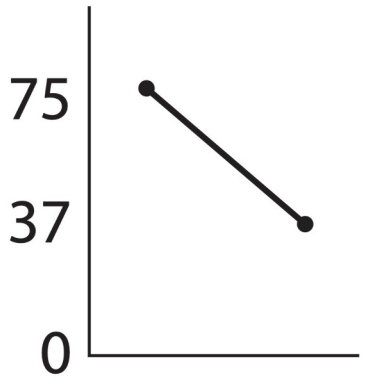
Bars



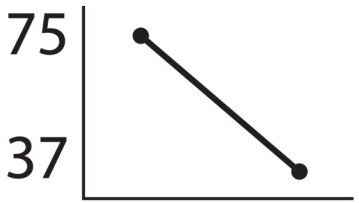


between two quantities

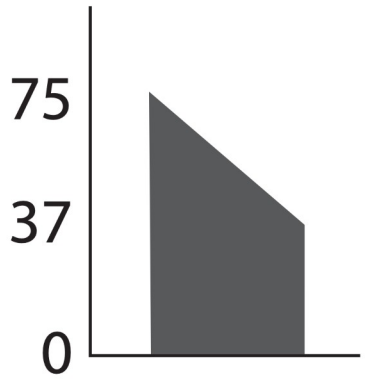
Line graphs



a



b



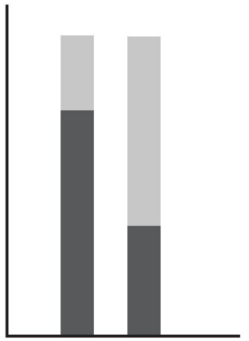
c

between two quantities

Percentage bars



a



b

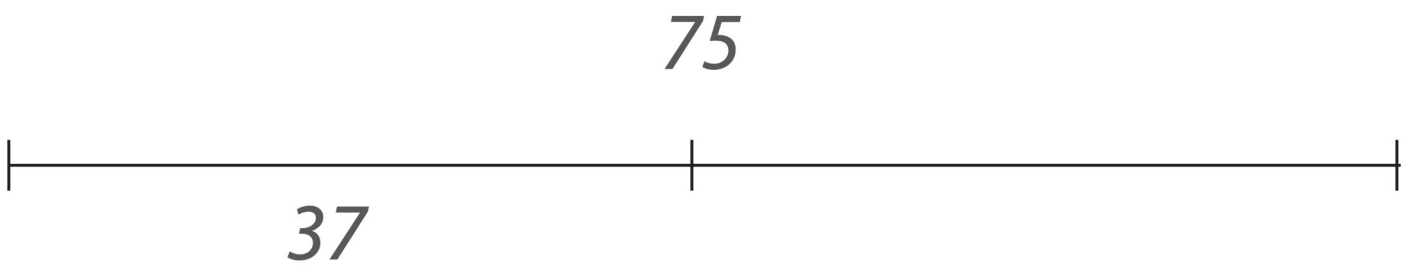
between two quantities

Spliced bar



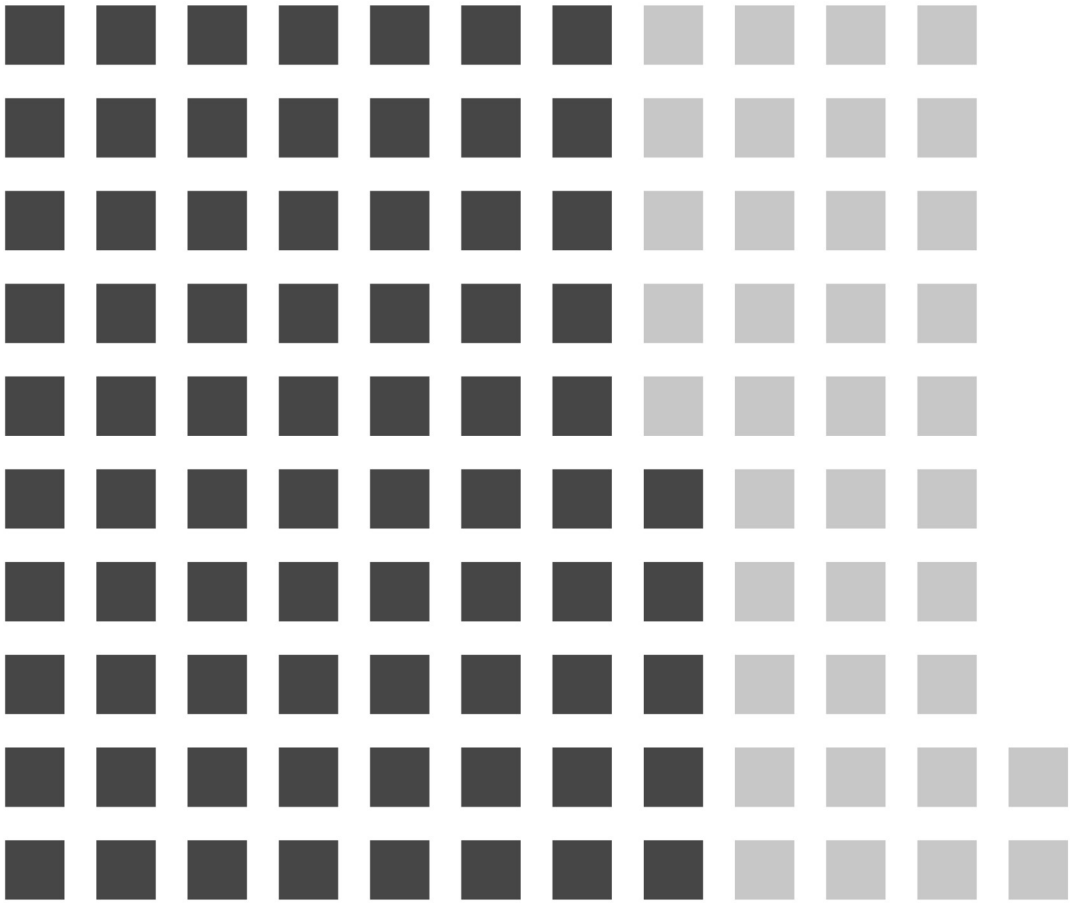
between two quantities

Porportion



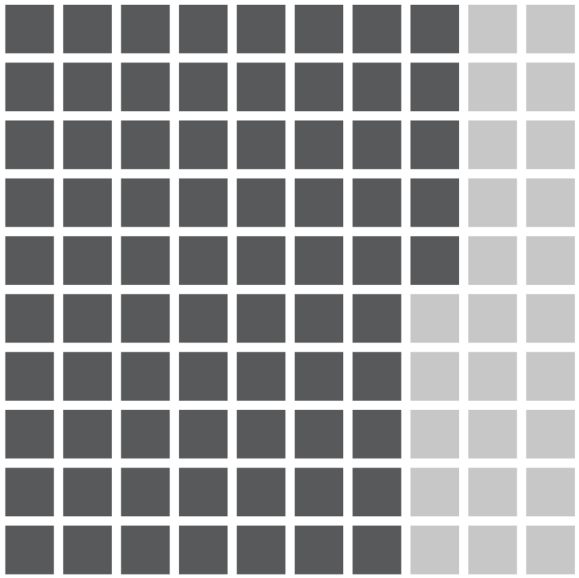
between two quantities

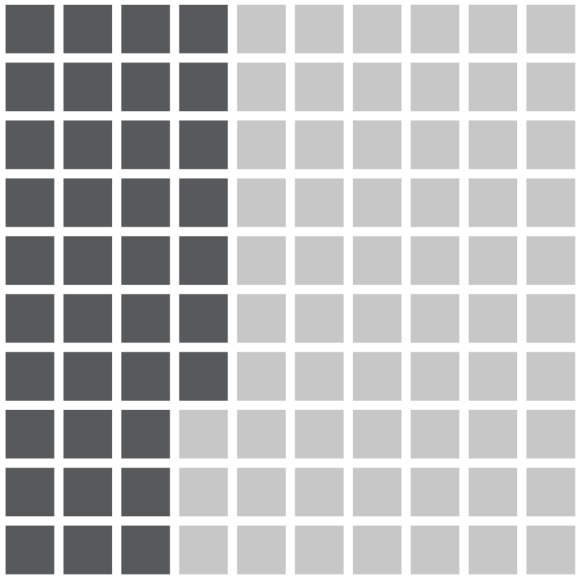
Merged Squares (Waffle Chart)



between two quantities

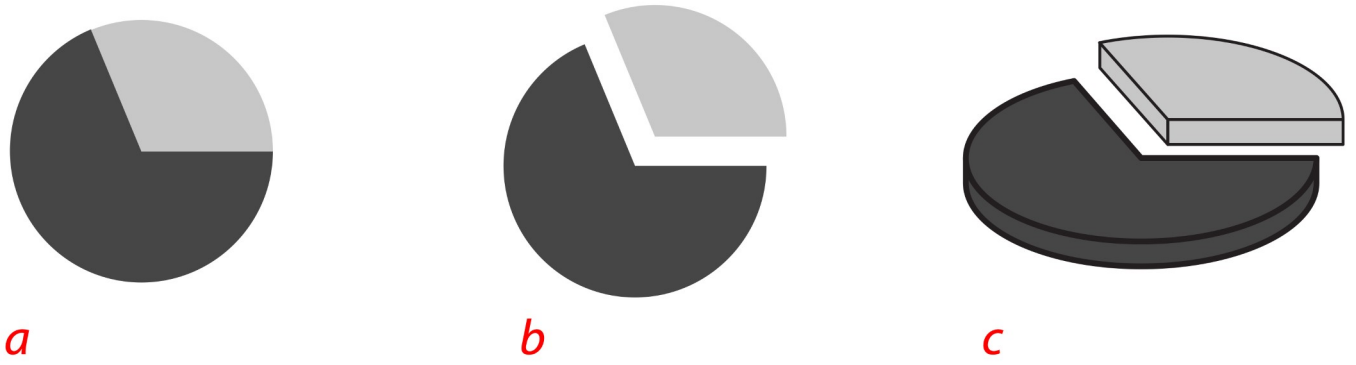
Percentage in Squares





between two quantities

Pie Charts



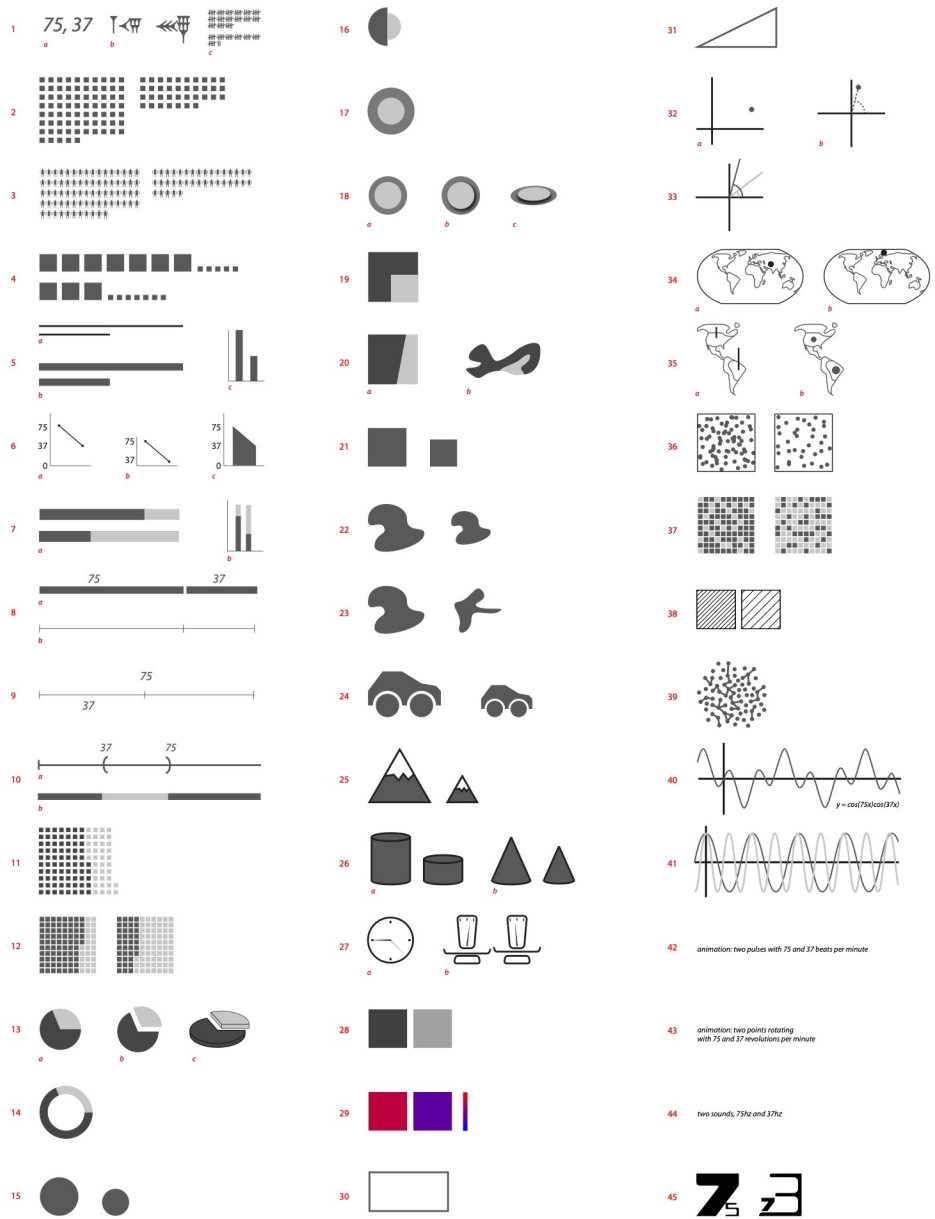
a *b* *c*

between two quantities

Circle Areas



75, 37 multiple ways to communicate two quantities



Santiago Ortiz 2012, from the post: ***http://blog.visual.ly/45-ways-to-communicate-two-quantities***

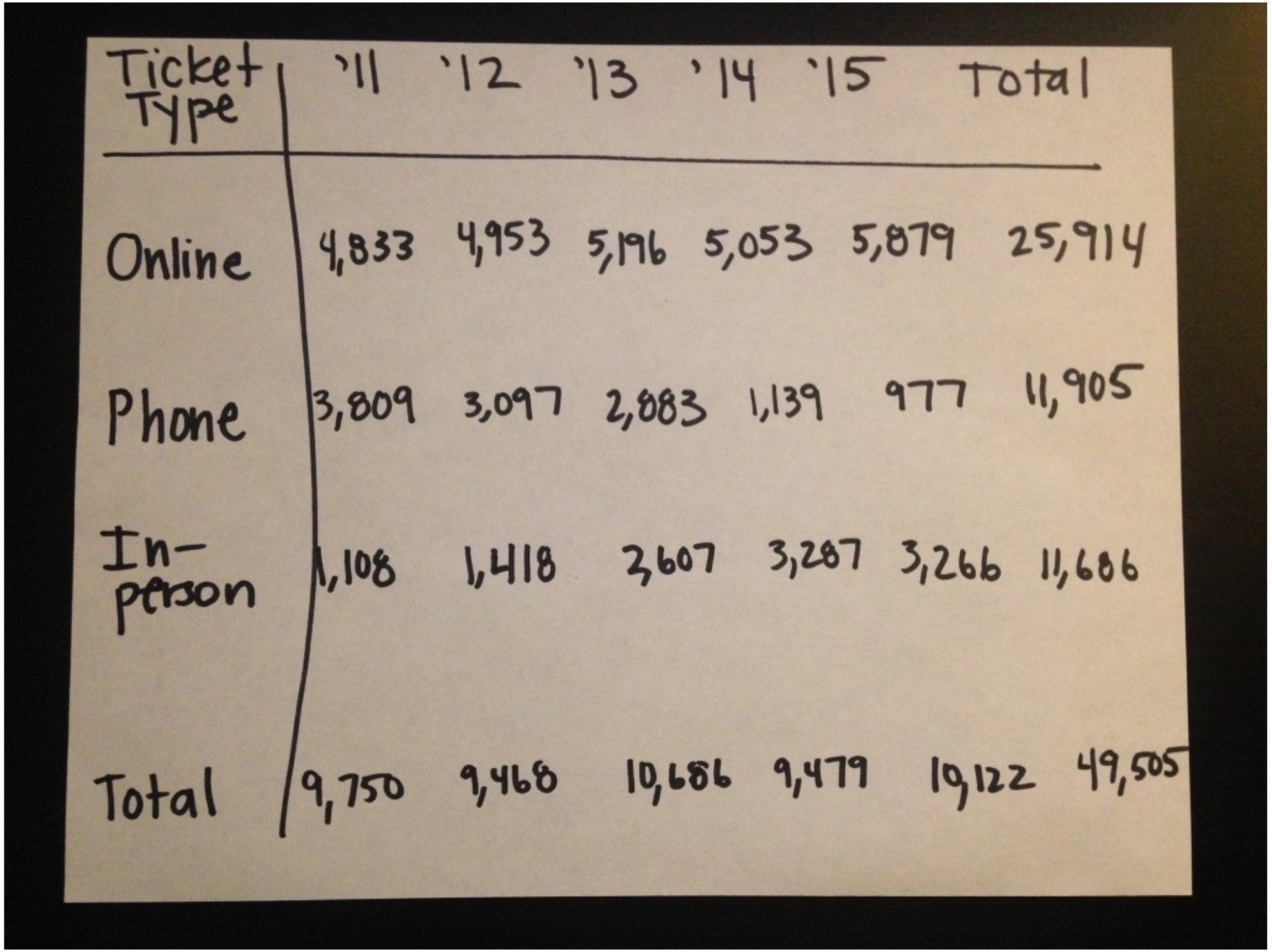
Sketch
CHARLOTTESVILLE

Some principles for sketching

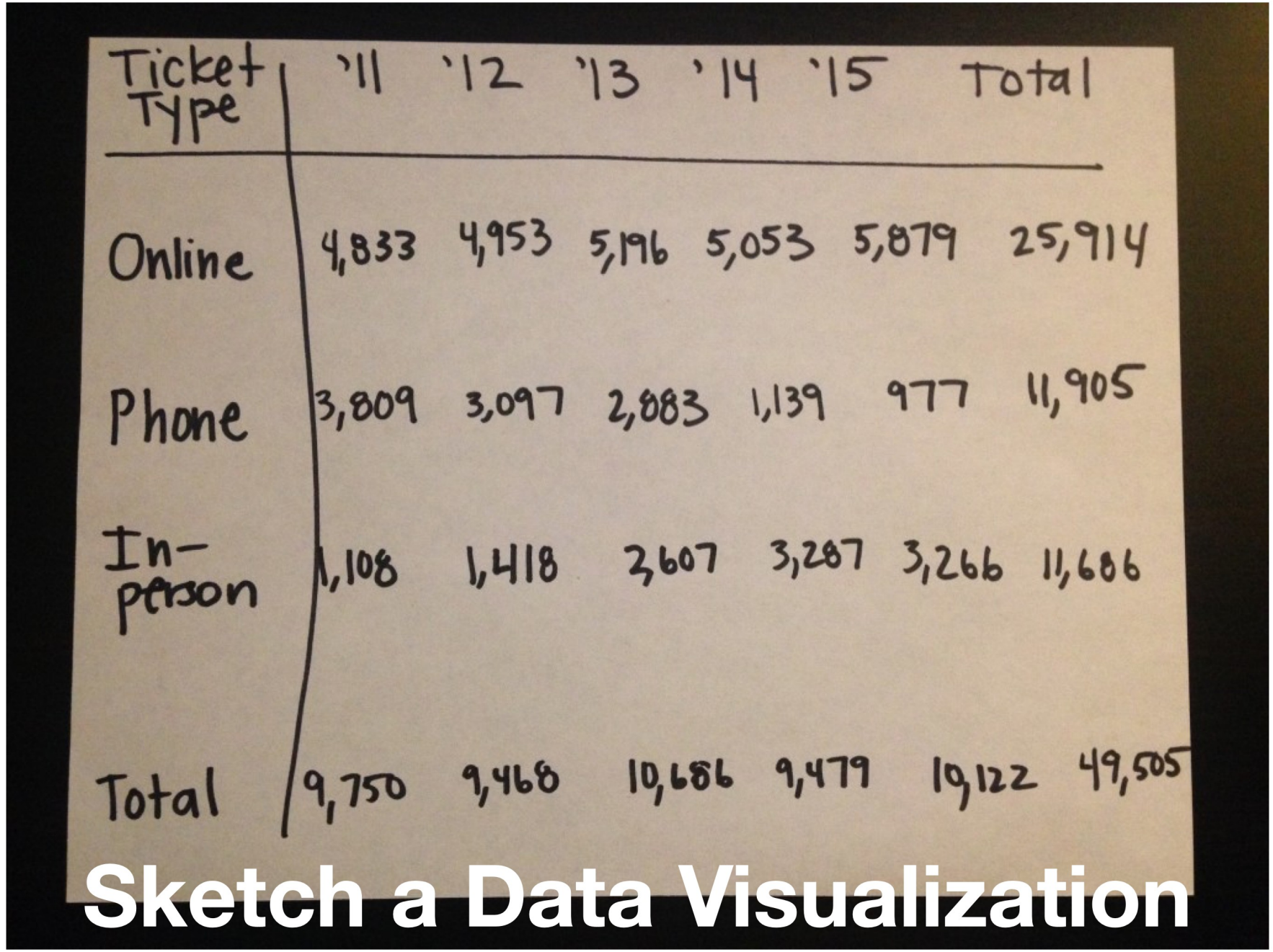
- use as few lines as you can
- communicate the essence of the idea
- details only if they are important
- choose the detail you put in deliberately

SKETCH SOME DATA

- What is the story here?
- What is the most interesting way to show the story?

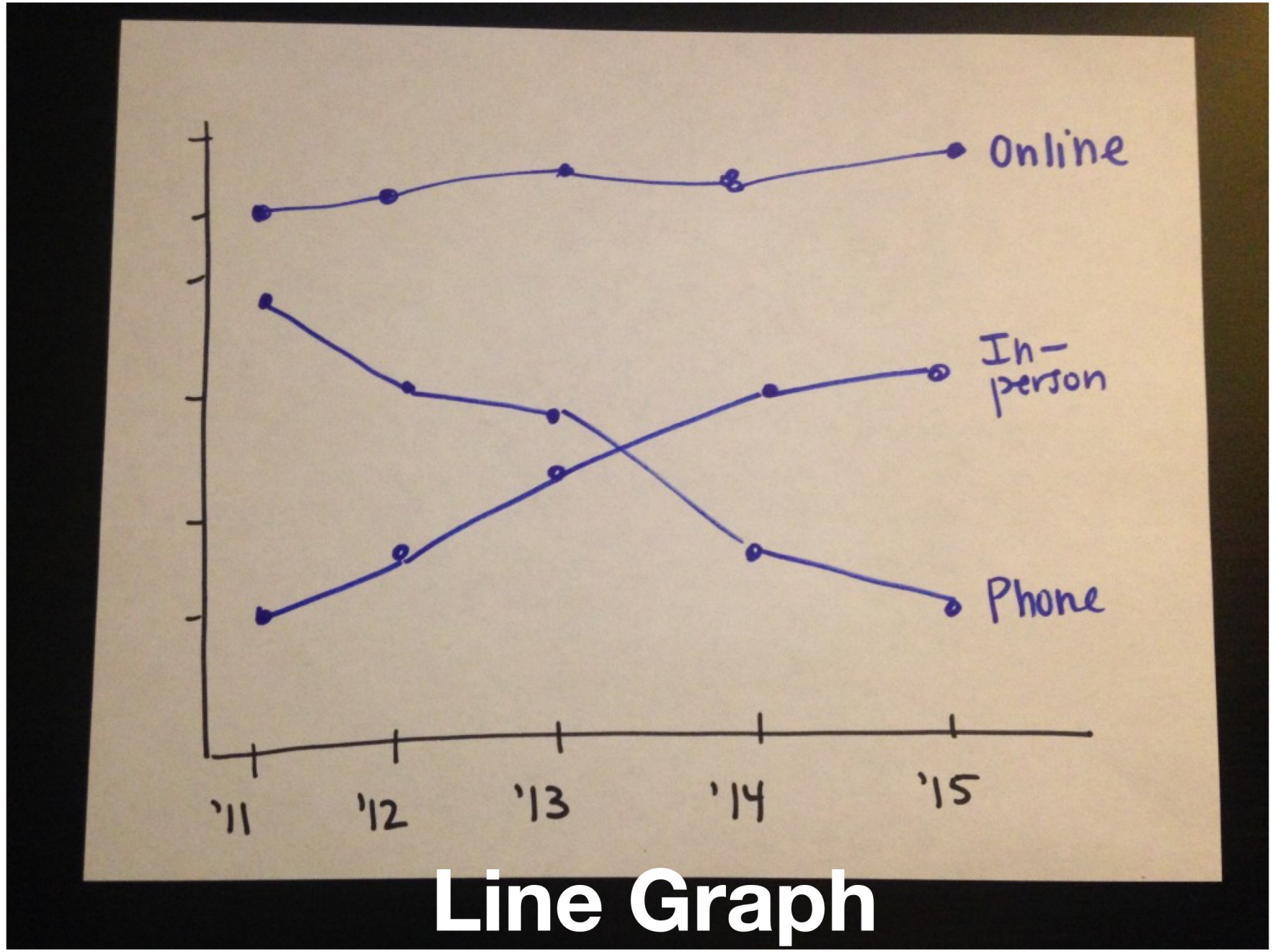


Ticket Type	'11	'12	'13	'14	'15	Total
Online	4,833	4,953	5,196	5,053	5,879	25,914
Phone	3,809	3,097	2,083	1,139	977	11,905
In-person	1,108	1,418	3,607	3,287	3,266	11,686
Total	9,750	9,468	10,686	9,479	19,122	49,505



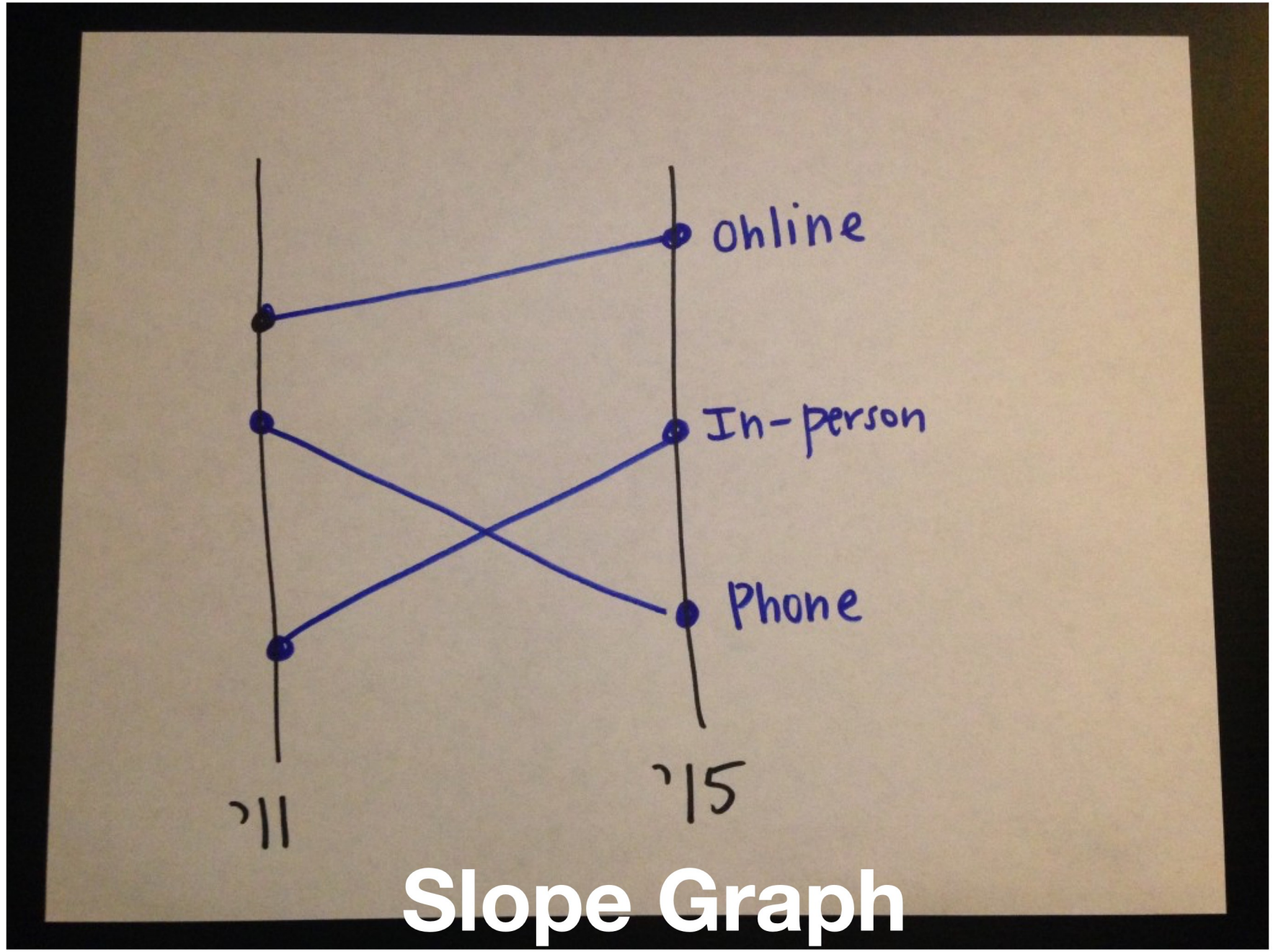
Sketch a Data Visualization

Credit: Ann Emery



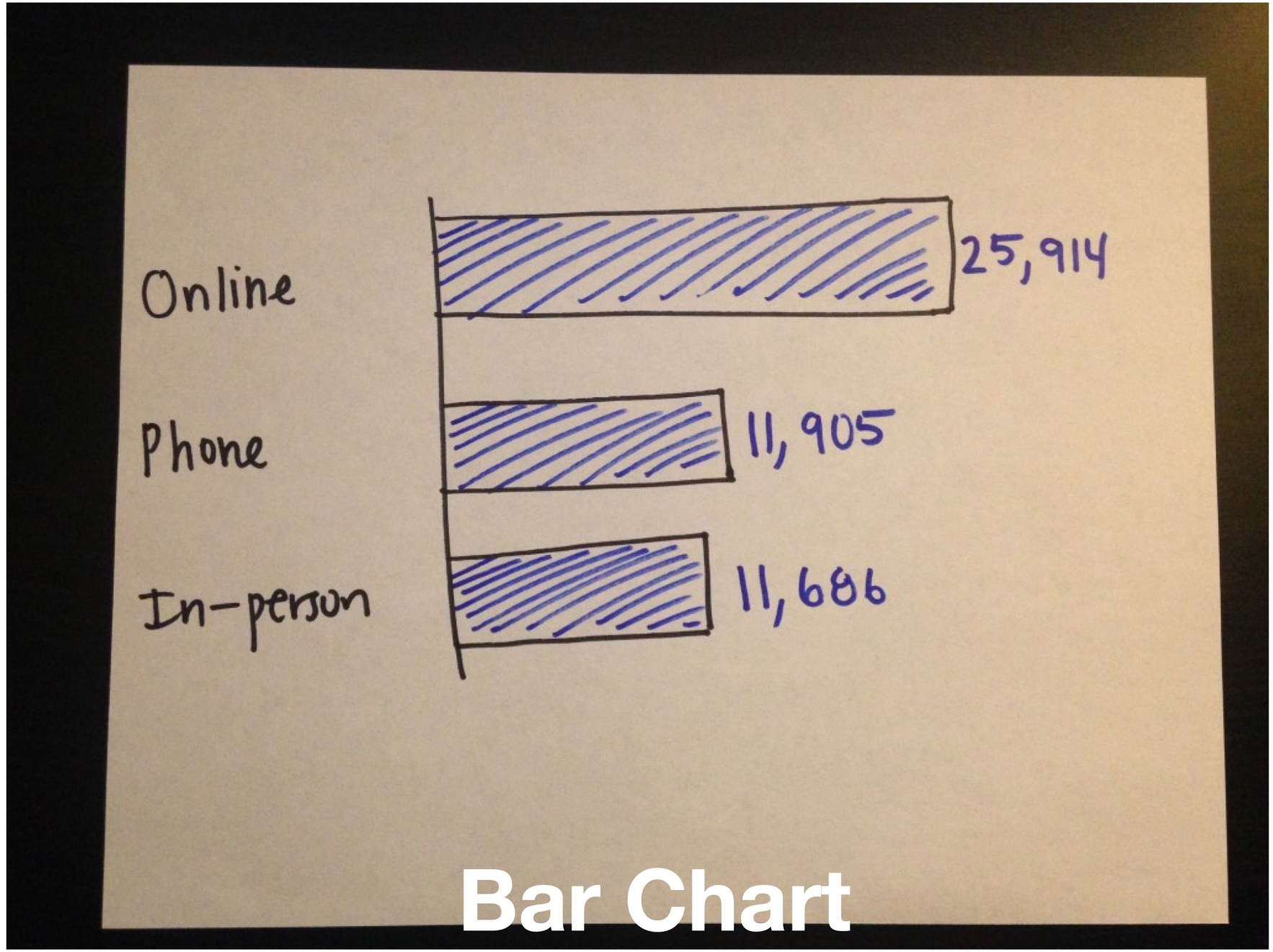
Line Graph

Credit: Ann Emery



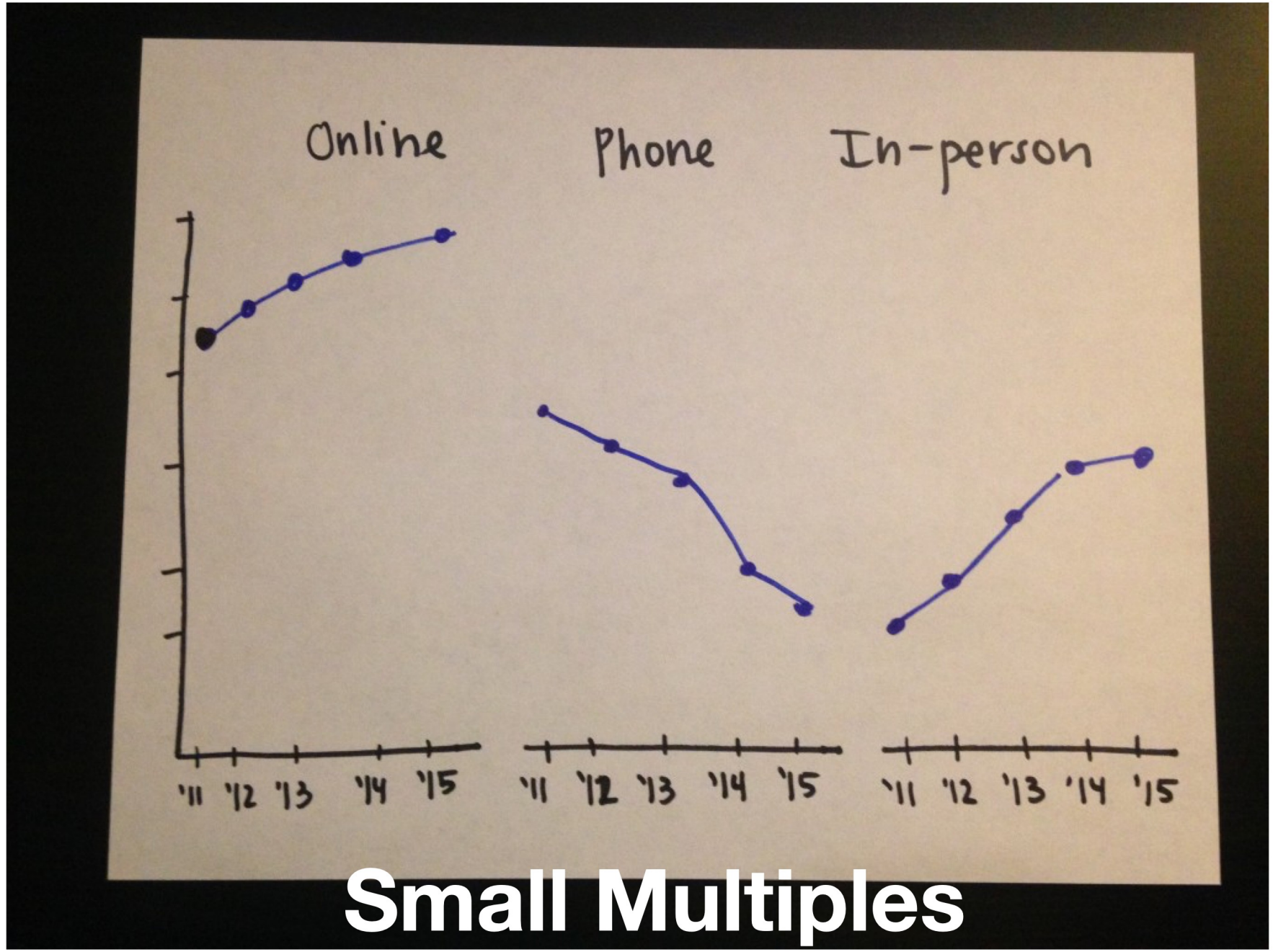
Slope Graph

Credit: Ann Emery



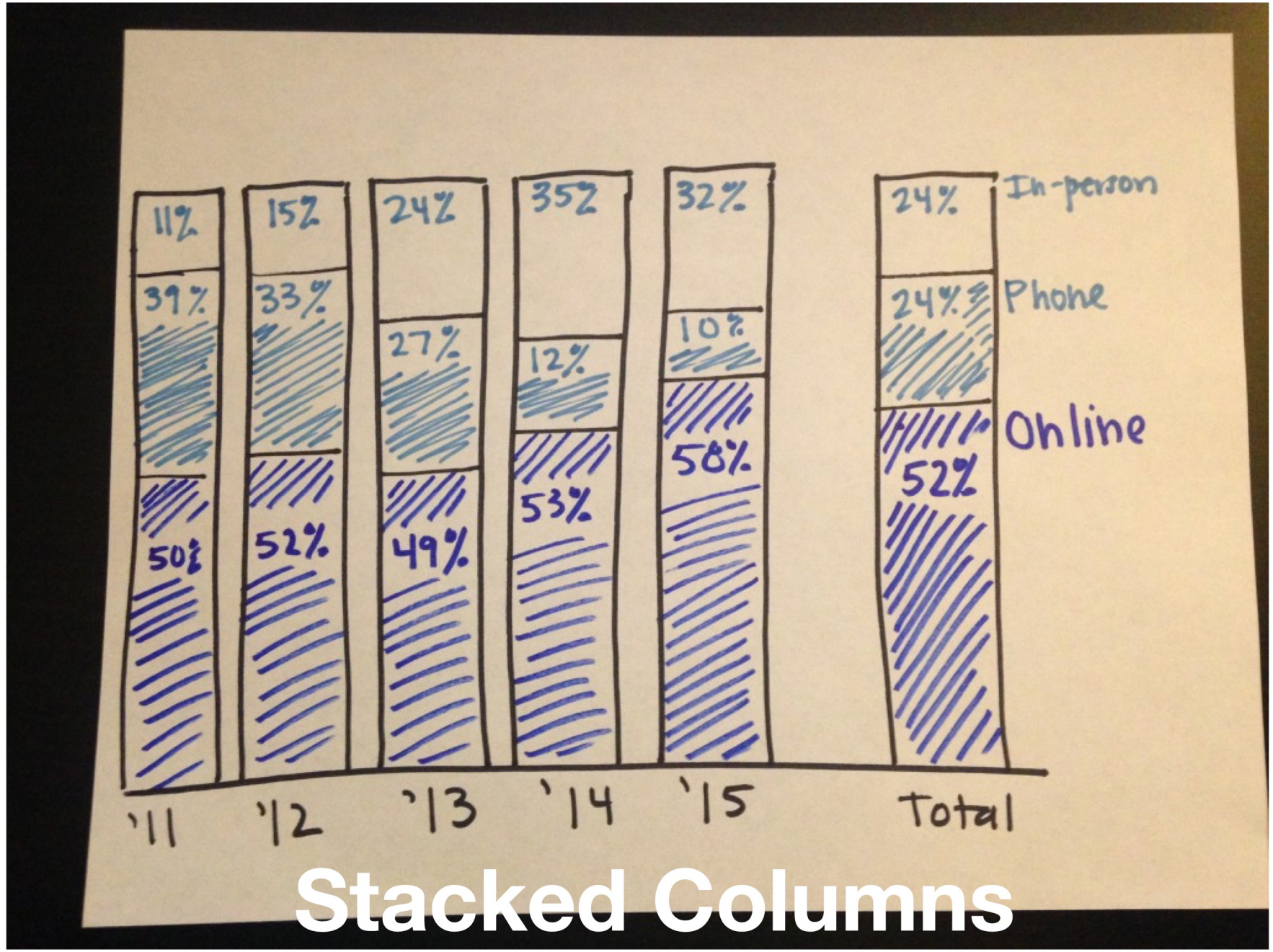
Bar Chart

Credit: Ann Emery



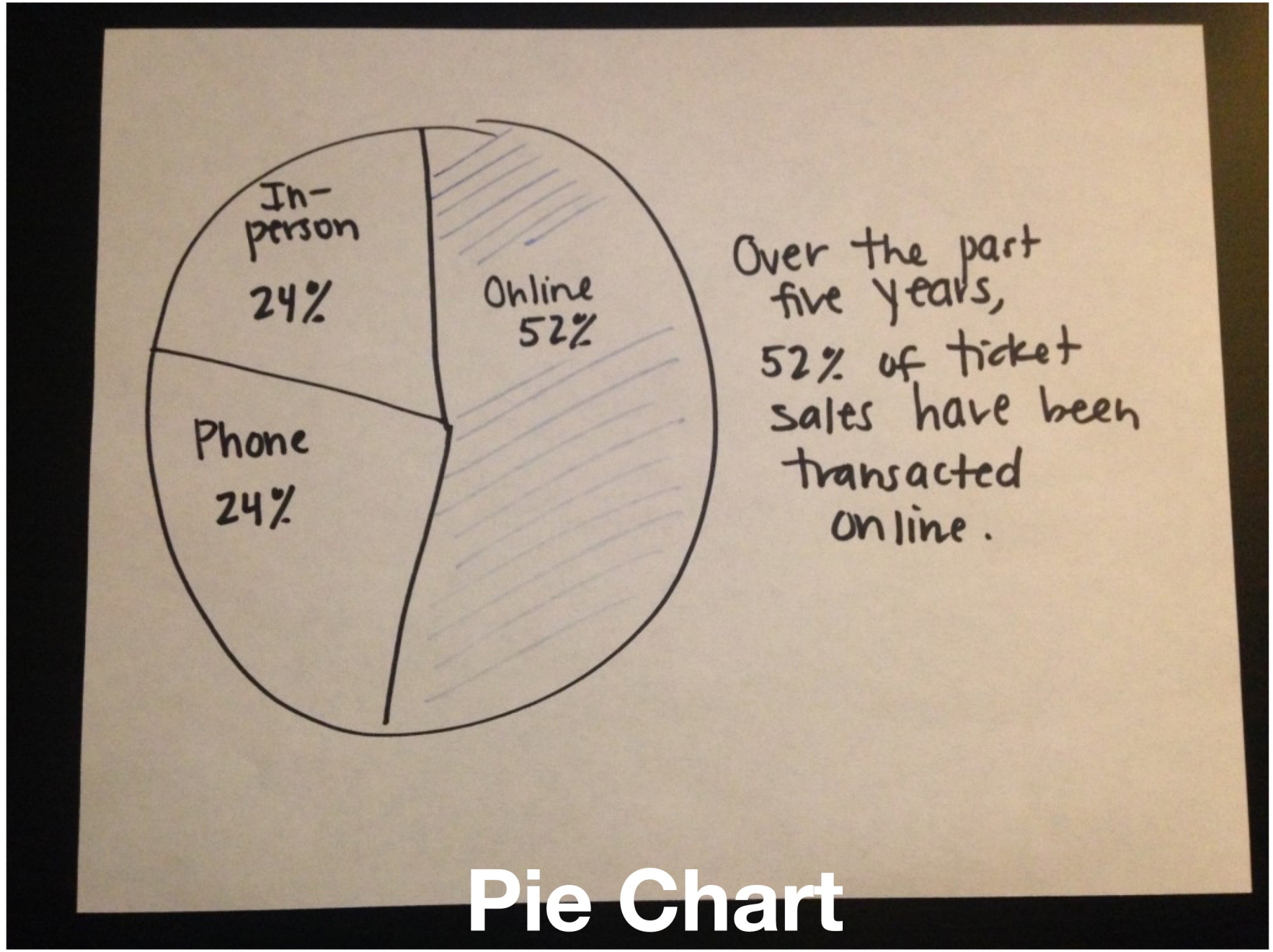
Small Multiples

Credit: Ann Emery



Stacked Columns

Credit: Ann Emery



Pie Chart

Credit: Ann Emery

Takeaways from infographics around SDS

- What would you implement in your own infographic or science vis?
- What would you avoid?

Sketching infographics

- Consider overall layout first; overview of information
  - Consider the story to tell
- Focus on data points– pick your encodings
- Prototype visualizations



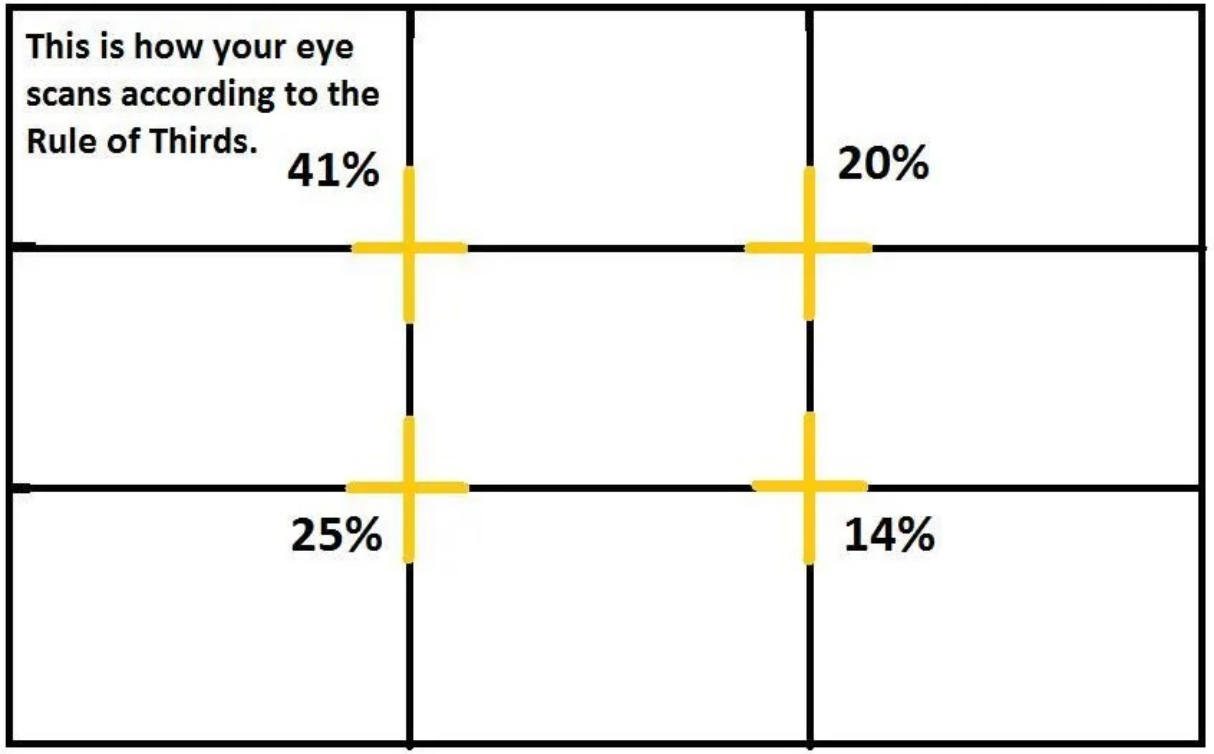
Lay out panels so the eye knows where to go

Most infographics are built from simple plot types: bar, line, pie.

Often organized like a document: read left to right, top to bottom.

Sometimes organized like a painting: one central focus with content around it.

Put the most important elements on the lines and intersections of a rule-of-thirds grid.



Where the eye lands on a rule-of-thirds grid

Rule-of-thirds image: “The rule of thirds in UX design,” Design Bootcamp (Medium)

In your group

1. **Read your story card and write a one-sentence lead** **3 min**

   Who is the audience? What do they need to believe or do? Write the single point a general should get in five seconds.

2. **Pick a format and map the parts to an infographic sketch** **7 min**

   Assign each story part to a panel. Name the chart type.
   Sketch the infographic, considering overall layout
   This is a QUICK sketch. You will have time to revise. Just get the main points down.

   TAKE A PHOTO AT THIS STAGE BEFORE REVISION.

3. **Swap and critique** **5 min**

   Write down any feedback that you find helpful.

4. **Revise** **10 min**

   Revise your sketch or make a new one.
   TAKE A PHOTO AT THIS STAGE AFTER REVISION.

Story 1: Four stars for everything

- Audience: moviegoers deciding what to see this weekend
- Data: fandango_score_comparison.csv: 146 rows × 22 columns, films from 2015 (FiveThirtyEight)
- One row is one film with 30+ fan reviews on Fandango, with its scores on Fandango, Rotten Tomatoes, Metacritic, and IMDb.

First five rows (7 of 22 columns)

FILM	RottenTomatoes	Metacritic	IMDB	Fandango_Stars	Fandango_Ratingvalue	Fandango_votes
Avengers: Age of Ultron (2015)	74	66	7.8	5	4.5	14846
Cinderella (2015)	85	67	7.1	5	4.5	12640
Ant-Man (2015)	80	64	7.8	5	4.5	12055
Do You Believe? (2015)	18	22	5.4	5	4.5	1793
Hot Tub Time Machine 2 (2015)	14	29	5.1	3.5	3	1021

Story 1: what the data shows

- The stars Fandango showed were higher than its own average rating for 130 of 146 films, and never lower
- It rounded up by as much as half a star: 13 films gained a full half star
- No film got fewer than 3 stars
- Average score out of 5: Fandango 4.09 stars; IMDb users 3.37; Rotten Tomatoes critics 3.04; Metacritic critics 2.94
- 73% of films got 4 stars or more on Fandango, against 7% on IMDb (scaled to 5)
- “Do You Believe?” got 5 stars on Fandango and 18% from Rotten Tomatoes critics
- “Paul Blart: Mall Cop 2” got 3.5 stars on Fandango and 5% from critics
- Context: Fandango makes its money selling movie tickets
- Uncertainty: fan vote counts range from 35 to 34,846 per film

Story 2: Birthdays are scheduled

- Audience: hospital administrators planning maternity ward staffing
- Data: US_births_2000-2014_SSA.csv: 5,479 rows × 5 columns, every day from 2000 to 2014 (Social Security Administration, via FiveThirtyEight)
- One row is one day in the U.S. and how many babies were born. day_of_week: 1 = Monday, 7 = Sunday.

First five rows (all 5 columns)

year	month	date_of_month	day_of_week	births
2000	1	1	6	9083
2000	1	2	7	8006
2000	1	3	1	11363
2000	1	4	2	13032
2000	1	5	3	12558

Story 2: what the data shows

- Weekends have 37% fewer births than weekdays (8,040 vs. 12,675 a day)
- Tuesday is the busiest day (13,122 on average); Sunday the quietest (7,518)
- Christmas Day has 41% fewer births than a typical day of the same weekday; New Year’s Day 28% fewer; July 4 22% fewer
- Halloween has 11% fewer births; Valentine’s Day has 3% more
- Friday the 13th has 5.1% fewer births than other Fridays of the same year
- Mid-September is busiest: September 12, 19, and 20 have the highest averages
- The busiest single day was 9/9/09: 16,081 births
- Births per year peaked at 4.38 million in 2007 and fell 9% by 2013
- Uncertainty: single Fridays swing about ±7%, but 23 of the 25 Friday the 13ths were below average

Story 3: Majors with more women pay less

- Audience: first-year students choosing a major
- Data: recent-grads.csv: 173 rows × 21 columns, American Community Survey 2010–2012 (via FiveThirtyEight)
- One row is one college major, with jobs and yearly earnings for its graduates under 28.

First five rows (8 of 21 columns)

Major	Major_category	Total	ShareWomen	Sample_size	Median	P25th	P75th
PETROLEUM ENGINEERING	Engineering	2339	0.120564344	36	110000	95000	125000
MINING AND MINERAL ENGINEERING	Engineering	756	0.101851852	7	75000	55000	90000
METALLURGICAL ENGINEERING	Engineering	856	0.153037383	3	73000	50000	105000
NAVAL ARCHITECTURE AND MARINE ENGINEERING	Engineering	1258	0.107313196	16	70000	43000	80000
CHEMICAL ENGINEERING	Engineering	32260	0.341630502	289	65000	50000	75000

Story 3: what the data shows

- Across 172 majors, the more women in a major, the lower its median pay (correlation −0.62)
- The 10 best-paid majors average 23% women; the 10 worst-paid average 79%
- Majors that are 75%+ women: median pay $32,400; majors with 25% or fewer women: $50,000
- Highest: Petroleum Engineering, $110,000, 12% women. Lowest: Library Science, $22,000, 88% women
- A typical engineering major pays $57,000 (24% women); education $32,750 (81% women)
- Pay varies within a major: the middle half of Computer Science grads earn $39,000–$70,000
- Uncertainty: 32 majors have fewer than 30 survey respondents; Library Science has 2
- Caveat: this compares majors, not men and women in the same job, and it doesn’t show why

Story 4: Smaller budgets, same returns

- Audience: studio executives deciding which films to fund
- Data: bechdel_movies.csv: 1,794 rows × 15 columns, films from 1970 to 2013 (BechdelTest.com and The-Numbers.com, via FiveThirtyEight)
- One row is one film, whether it passes the Bechdel test (two women who talk to each other about something besides a man), and its budget and box office. Columns ending in _2013$ are adjusted for inflation.

First five rows (7 of 15 columns)

year	title	clean_test	binary	budget_2013$	domgross_2013$	intgross_2013$
2013	21 & Over	notalk	FAIL	13000000	25682380	42195766
2012	Dredd 3D	ok	PASS	45658735	13611086	41467257
2013	12 Years a Slave	notalk	FAIL	20000000	53107035	158607035
2013	2 Guns	notalk	FAIL	61000000	75612460	132493015
2013	42	men	FAIL	40000000	95020213	95020213

Story 4: what the data shows

- 45% of films pass (803 of 1,794)
- Why films fail: in 514 the women never talk to each other, in 194 they only talk about men, and 141 have fewer than two women
- Passing films get smaller budgets: median $31.5M vs. $44.0M for films that fail (2013 dollars)
- Only 31% of $100M+ films pass, compared with 52% of films under $20M
- Per dollar spent they earn about the same: median worldwide gross is 2.7× the budget for passing films, 2.6× for failing ones
- The pass rate rose from 26% in the 1970s to 49% in the 2000s
- Users disputed 425 of the ratings (the “-disagree” tags in the test column)
- Watch for errors: Shaft (1971) is listed with a $305M budget, likely the 2000 remake’s numbers
- Uncertainty: returns vary hugely (the middle half runs from about 1.3× to 5×), so “earn the same per dollar” holds up but “earn more” doesn’t

Story 5: Past crashes don’t predict future crashes

- Audience: travelers nervous about choosing an airline
- Data: airline-safety.csv: 56 rows × 8 columns, 1985–1999 and 2000–2014 (Aviation Safety Network, via FiveThirtyEight)
- One row is one airline, with its incidents, fatal accidents, and deaths in two 15-year periods. An asterisk means regional subsidiaries are included.

First five rows (all 8 columns)

airline	avail_seat_km_per_week	incidents_85_99	fatal_accidents_85_99	fatalities_85_99	incidents_00_14	fatal_accidents_00_14	fatalities_00_14
Aer Lingus	320906734	2	0	0	0	0	0
Aeroflot*	1197672318	76	14	128	6	1	88
Aerolineas Argentinas	385803648	6	0	0	1	0	0
Aeromexico*	596871813	3	1	64	5	0	0
Air Canada	1865253802	2	0	0	2	0	0

Story 5: what the data shows

- Flying got safer: fatal accidents fell from 122 to 37 (−70%), and deaths from 6,295 to 3,109
- 32 of 56 airlines had no fatal accidents from 2000 to 2014
- An airline’s incidents in 1985–99 somewhat predict its incidents in 2000–14 (correlation 0.4)
- Its deaths don’t predict later deaths at all (correlation 0.05)
- Japan Airlines (520 deaths) and Korean Air (425) were among the worst in 1985–99; neither had any deaths in 2000–14
- 5 airlines with no fatal accidents in 1985–99 had at least one later, including Alaska and SAS
- Malaysia Airlines had the most deaths in 2000–14: 537, from 2 crashes
- Bigger airlines fly more, so compare per seat-kilometer, not raw counts
- Uncertainty: fatal crashes are rare, so one or two crashes swing an airline’s record

Swap and critique

The generals
Can someone get your lead in five seconds?

Roles
Does each panel play one clear role in the arc?

Order
Does raw data come before derived numbers?

Variety
Do different analyses use different chart types?

Focus
Is anything on the page tangential to the story?

Honesty
Does the story claim more than the data shows?