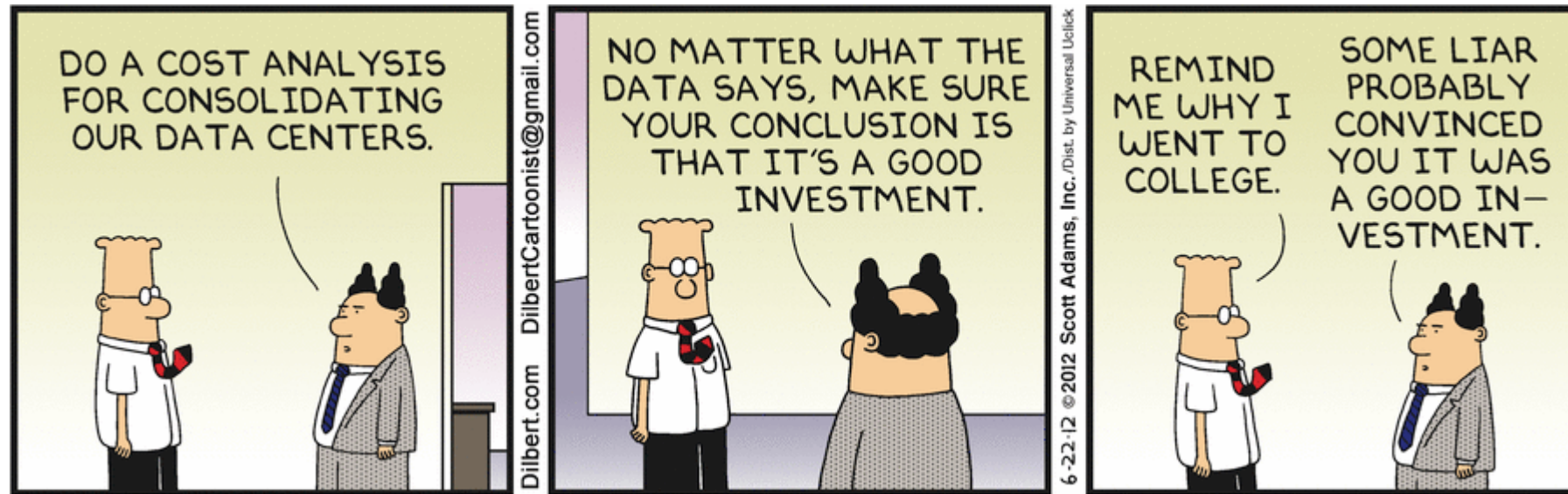


Introduction to Data Analysis

Ori Plonsky

Spring 2023



Agenda

- Data science
 - What?
 - Why?
 - Who?
 - How?
- Exercise
- Logistics

Data Science: What?

An interdisciplinary field that combines domain expertise, programming skills, and statistical knowledge to extract insights and knowledge from data.

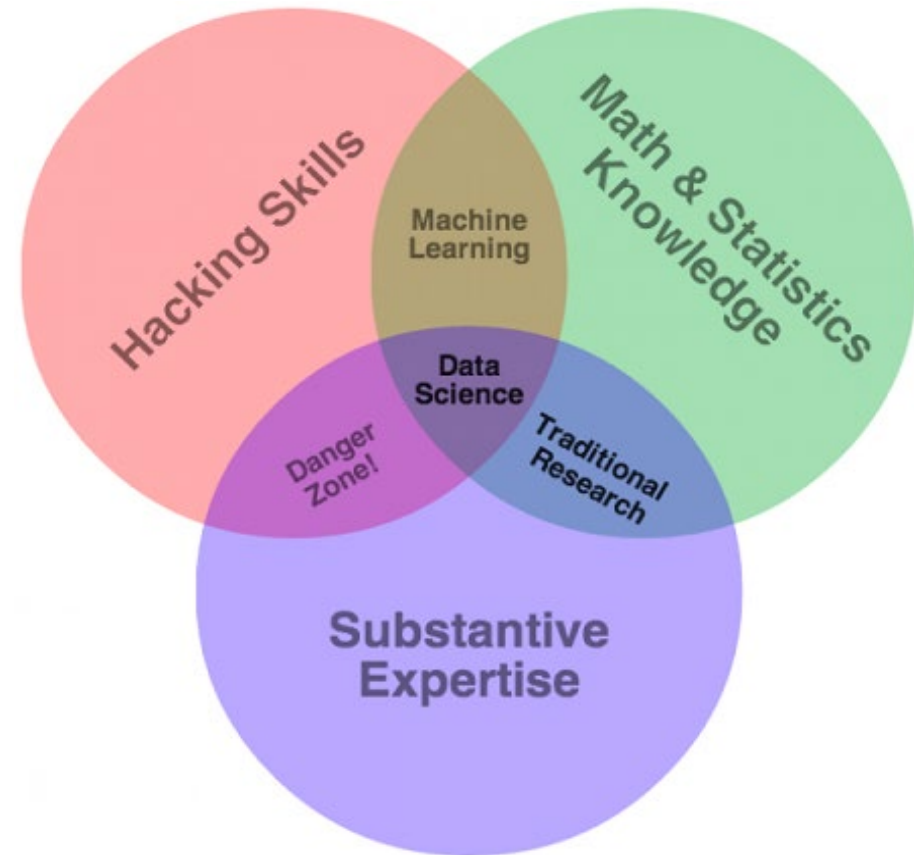


Image taken from Harvard CS109

Data science: What?

Extracting useful insights from data using computation, statistics, and visualization

Data science: What?

Extracting useful insights from data using computation, statistics, and visualization

- **Exploration**
 - Play with the data, let it guide the questions
 - Identify patterns in information
 - Uses visualizations

Data science: What?

Extracting useful insights from data using computation, statistics, and visualization

- **Exploration**

- Play with the data, let it guide the questions
- Identify patterns in information
- Uses visualizations

- **Inference**

- Start with a formal question
- Quantify whether patterns are reliable
- Uses randomization and statistical decision theory

Data science: What?

Extracting useful insights from data using computation, statistics, and visualization

- **Exploration**
 - Play with the data, let it guide the questions
 - Identify patterns in information
 - Uses visualizations
- **Inference**
 - Start with a formal question
 - Quantify whether patterns are reliable
 - Uses randomization and statistical decision theory
- **Prediction**
 - Make informed guesses
 - Uses machine learning

Data science: What?

Extracting useful insights from data using computation, statistics, and visualization

- **Exploration**

- Play with the data, let it guide the questions
- Identify patterns in information
- Uses visualizations

- **Inference**

- Start with a formal question
- Quantify whether patterns are reliable
- Uses randomization and statistical decision theory

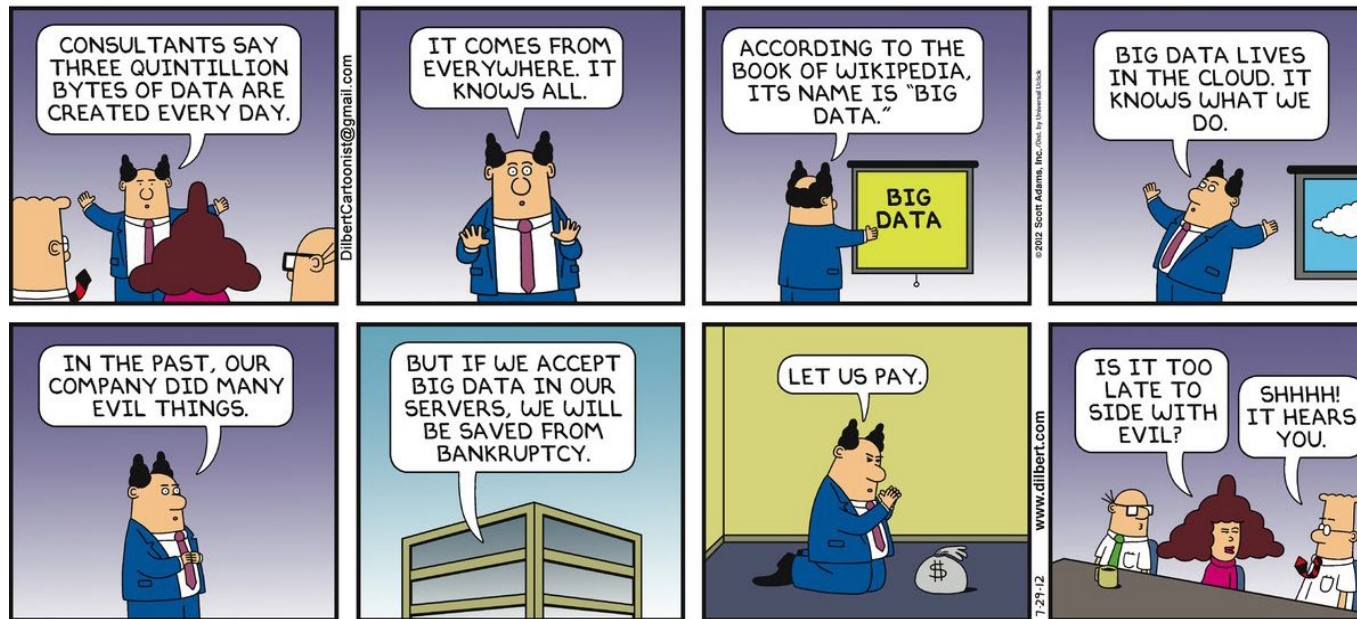
- **Prediction**

- Make informed guesses
- Uses machine learning



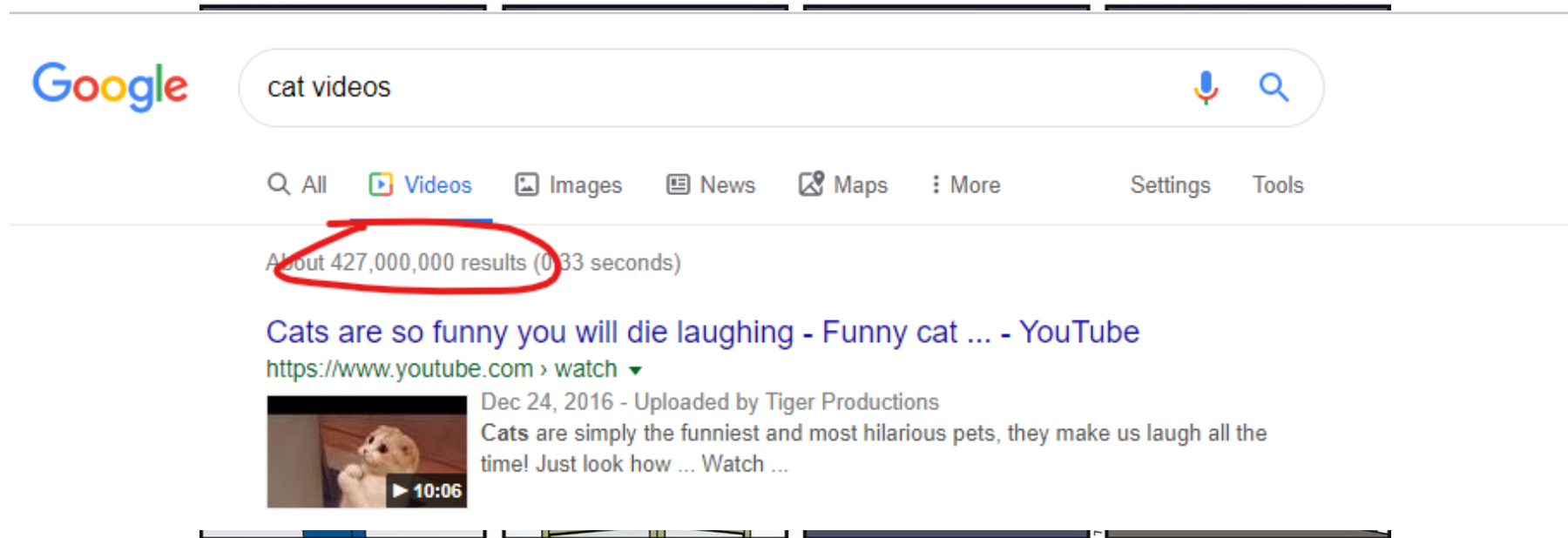
Data Science: Why?

Lots of data...



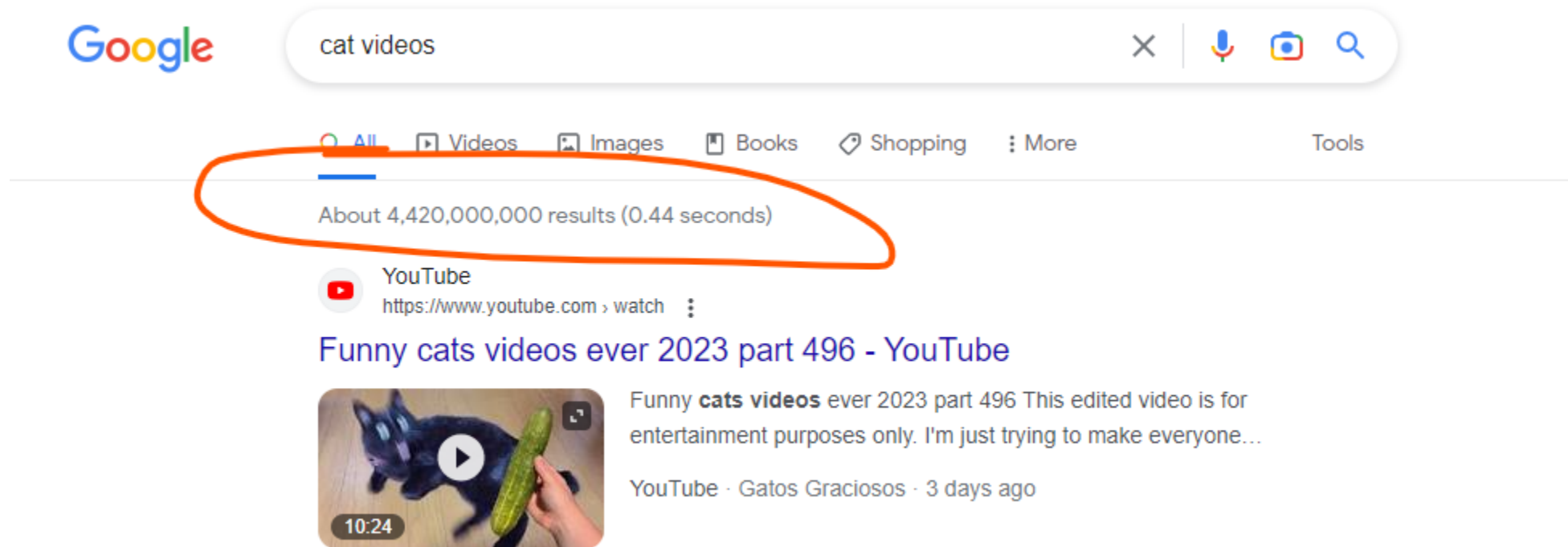
Data Science: Why?

Lots of data...



Data Science: Why?

Lots of data...



The image shows a Google search interface. The search bar contains the text "cat videos". Below the search bar, the "All" tab is selected and highlighted with a blue underline. A red hand-drawn oval encircles the text "About 4,420,000,000 results (0.44 seconds)". Below this, the first search result is from YouTube, titled "Funny cats videos ever 2023 part 496 - YouTube". The result includes a video thumbnail showing a black cat and a hand holding a cucumber, with a play button in the center. The video duration "10:24" is visible in the bottom left corner of the thumbnail. To the right of the thumbnail, the text reads: "Funny **cats videos** ever 2023 part 496 This edited video is for entertainment purposes only. I'm just trying to make everyone..." and "YouTube · Gatos Graciosos · 3 days ago".

Google

cat videos

All Videos Images Books Shopping More Tools

About 4,420,000,000 results (0.44 seconds)

YouTube
[https://www.youtube.com > watch](https://www.youtube.com/watch)

Funny cats videos ever 2023 part 496 - YouTube

10:24

Funny **cats videos** ever 2023 part 496 This edited video is for entertainment purposes only. I'm just trying to make everyone...

YouTube · Gatos Graciosos · 3 days ago

Data Science: Why?

- “The ability to take **data** – to be able to **understand** it, to **process** it, to **extract value** from it, to **visualize** it, to **communicate** it – is going to be a hugely important skill in the next decades, not only at the professional level but even at the educational level for elementary school kids, for high school kids, for college kids. Because now we really do have essentially free and **ubiquitous data**.”

– Hal Varian, Chief Economist, Google

Data Science: Why?

5 Hottest Tech Jobs To Go For in 2023

ICT system developers, data analysts, and engineers are the future of data-driven businesses.



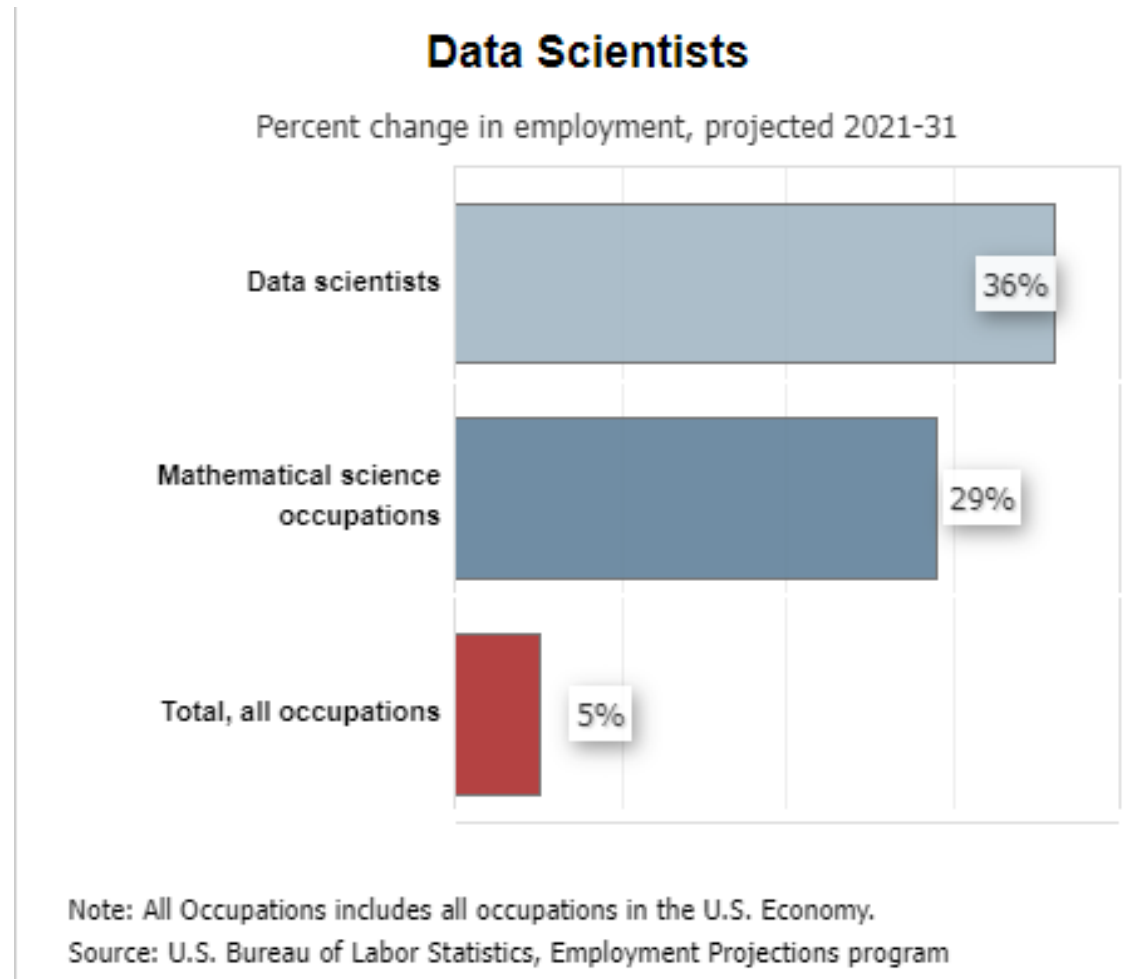
Paul Lewis CEO & Founder, Calamu

March 20, 2023

4. Data Analyst

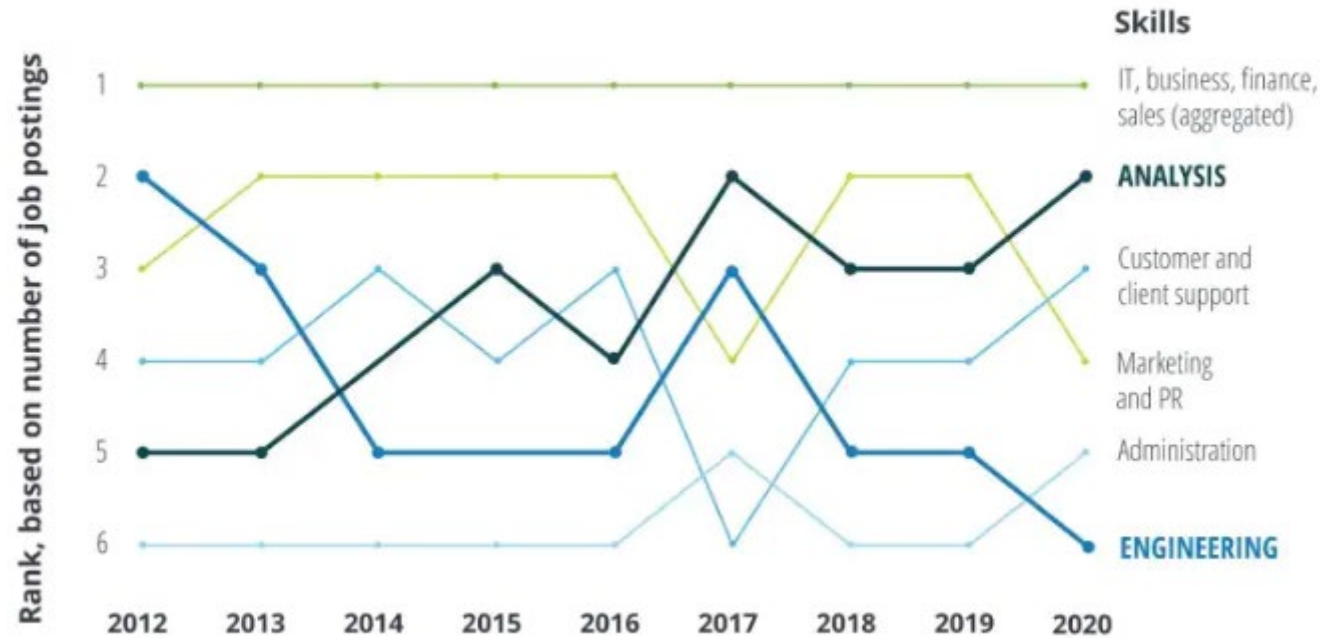
As data is becoming the modern currency for businesses, the role of a data analyst is unquestionably one of the most coveted jobs in 2023. There were 14,093 jobs up for grabs in January, a +41% year-on-year increase. Similar roles such as data engineer (12,948 jobs) and data scientists (11,365 jobs) are also in top demand. These data-centric roles are all very financially rewarding, with advertised salaries ranging from \$87,227 to \$148,616.

Data Science: Why?



Data Science: Why?

The US tech industry is looking for more 'analytical' and fewer 'engineering' talent



Note: At rank #1, we have combined/aggregated job posting numbers for four skills: IT, business, sales, and finance.
Source: Deloitte analysis based on data from the Burning Glass database.

Data Science: Who?

“A data scientist is someone who knows more statistics than a computer scientist and more computer science than a statistician.”

- Josh Blumenstock, Berkley

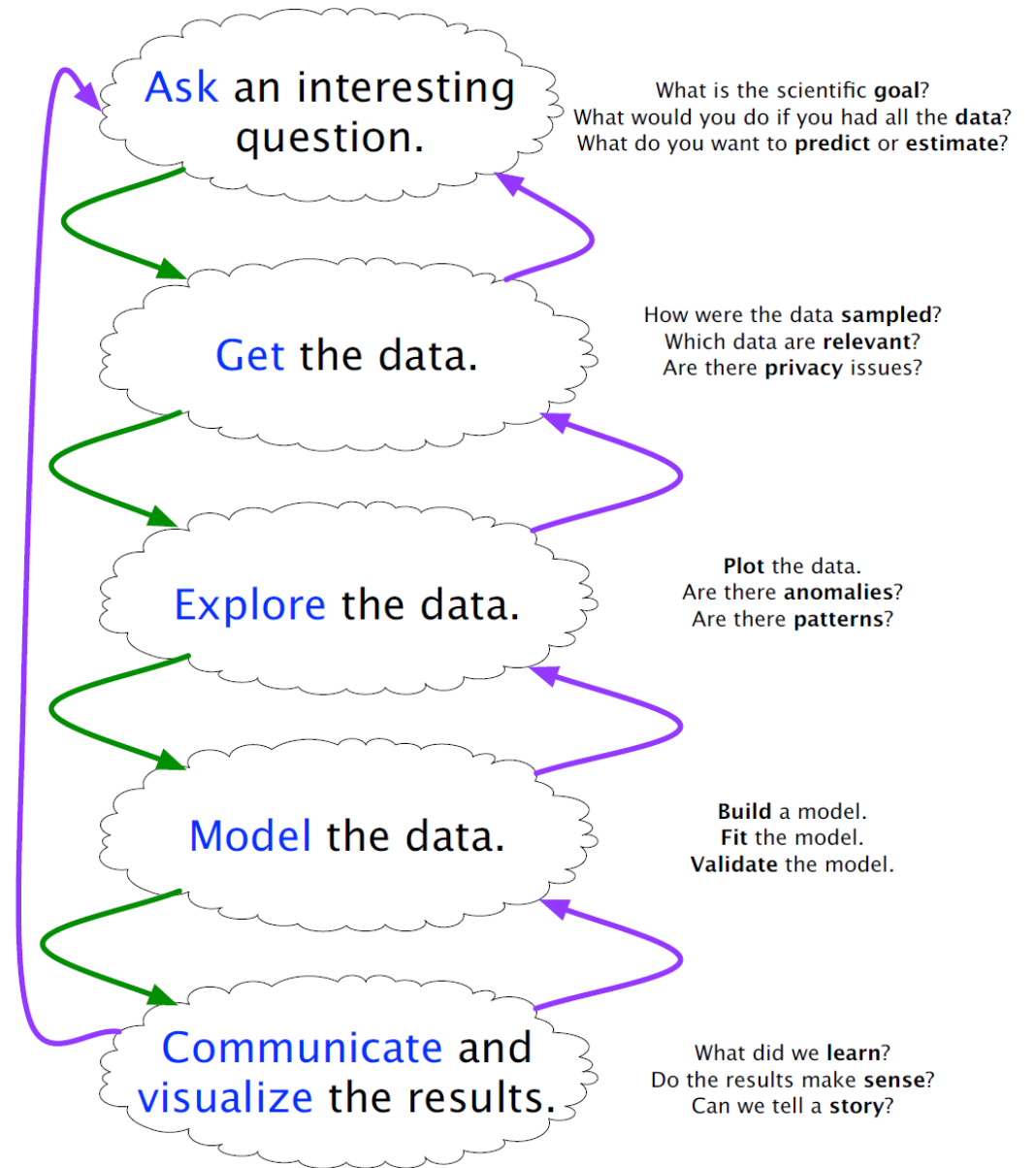
“Data Scientist = statistician + programmer + coach + storyteller + artist”

- Shlomo Aragam, Illinois Tech

Data; Science

- Data?
 - Name types of data
- Science?
 - As in a way of thinking

Data Science: How?



Example: Gender Tropes

- Question:
Does our culture, as reflected in movies, promotes certain gender roles?

Example: Gender Tropes

- Question:
Does our culture, as reflected in movies, promotes certain gender roles?
- Data: 2000 movies' screen direction

DIALOGUE



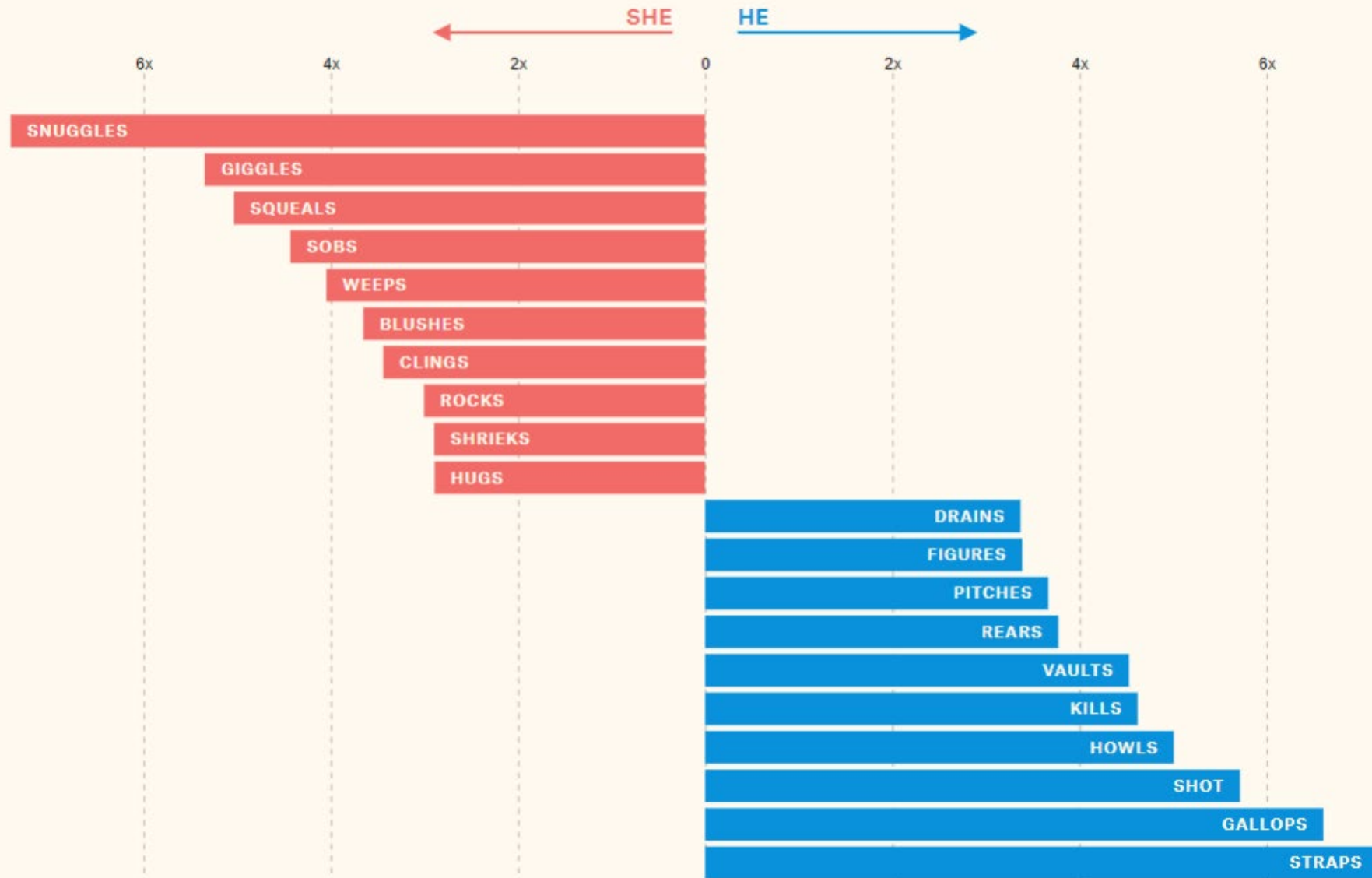
SCREEN DIRECTION

Rose gasps. There is nothing in her field of vision but water...She leans forward, arching her back. He puts his hands on her waist to steady her.

Rose closes her eyes...she smiles dreamily, then leans back, gently pressing her back against his chest. He pushes forward slightly against her.

The most used words for women vs. men

Likelihood that certain words appear after "she" vs. "he" in screen direction.



EVEN
MORE "SHE"  MORE "HE"

snuggles giggles squeals sobs weeps blushes clings rocks shrieks hugs shrinks gasps responds trembles pets flinches arches skips utters shudders startles buries swats murmurs resists
hovers caresses awakens shivers screams dances beats absently flees cleans stirs straddles cries moans bites realises mouths accepts wore smiles laughs wrote serves scoots liked
arranges scampers storms twirls softens ignores softly faints wonders fades sags hesitates casts applies hisses fiddles kisses sings awkwardly smokes stretches sips unbuttons stiffens
hurriedly hurries dries look locates threw smooths nods stubs instantly reappears wraps clearly refuses tiptoes lingers beams pivots curls glides strokes meant abruptly retrieves bursts
rakes relents reluctantly frantically understands fidgets breaks recoils hums looked pretends trails frowns retreats gonna licks touches reacts nearly sighs backs embraces squirms panics
yelps ends allows flashes senses exits gathers collects receives drapes folds undoes answers shakes instinctively replies told freezes resumes creeps calms gives rushes sails tentatively
thrashes becomes types avoids clips struggles fight considers leaves thumbs cradles covers blinks dabs meets rests regards tilts attacks darts eyes brushes descends gently nervously
returns forces closes fixes inhales flings scoops desperately plunges appears glimpses clicks halts wakes bolts slaps fumbles quickly wears clasps faces feeds barely shrugs believes
floats left whacks blows immediately emerges seems slips almost stares tugs happily hates slices runs leans sounds washes swallows cranes observes accidentally marches rifles
squeezes begins kneels turns leads snarls rolls wipes whirls tell flushes peeks shuts sits grimaces frees put calls glares tucks like plops scurries whispers tries remains actually
continues pinches tells lets yawns disappears heads looks chokes discovers plugs springs watches cautiously opens clutches studies dropped massages obeys suddenly loves scowls
crosses packs scribbles spits sneaks puts likes just lashes topples hangs lies angles starts claps adds flicks slowly cups angrily really stops pours jabs traces unzips crawls died
grasps slugs steadies breathes glances pushes directs inches hands reads comes unfolds winds attempts rub snorts walks drifts sinks hides goes swivels feels keeps sprays sways
means ducks races repeats used steps sends finishes talking trips locks waits gets snatches chooses obviously takes decides plucks slides moves peers holds claws already stomps
swipes asks admires met said stands buttons owns says sets strips must wanted focuses fell unwraps edges unlocks remembers shines reaches jumps always places climbs stumbles
handles leafs needed passed surreptitiously concentrates helps interrupts reels scrambles steels blocks clenches floors gags splashes rips notices knocks enters finally rises listens
quietly jerks bumps wants lays kicks flies makes picks throws casually scans winces might dangles hefts flails waves dresses realizes passes catches plants thinks knows rings empties
hears sweeps may signs wrenches swims shares started recovers hops props tightens indicates hear dashes got peels will silently probably grows whips finds spins pulls switches
lights assumes ties digs hopes wheels lines performs settles tenses sniffs stabs wanders downs pounds notes slams twists shows weaves bends bounces curses expects hastily pokes
can sees acts scrolls brings arrives needs follows get zips manages proceeds deposits hardly strikes exhales still smells came tears yanks lifts knew shifts presses grabs excuses
straightens hustles speeds recognizes carries pays also falls stays tastes drags never speaks dials turned slows relaxes brought dumps sticks changes know come lowers flips bought
sleeps greet registers succeeds now even withdraws cracks writhes saw nudges scratches hobbles steals pauses collapses offers puffs grinds braces thought expected levels gave hits
drops spots lurches clambers ever eventually snaps writes searches gazes approaches selects eases dips cocks groans lives works winks swerves grips fills rummages loved joins
regains wiggles talks beckons gulps maneuvers zooms stalks seizes vanishes points thrusts hauls leaps hit adjusts heard shoots refers deals honks rams releases clears nears pries
hurls mutters extends sucks trudges found removes accelerates orders produces inspects tumbles cuts calmly presents scrapes crushes pops checks drinks spies drives punches doubles
jams slings smirks fingers slumps prepares wrestles plays counts coughs pats gestures raises tosses lived stoops say contemplates hoists examines rap unscrews uses draws squats
shoves sprints rattles controls hooks surfaces loses delivers lunges dodges addresses called consults eats falters forgot swings shouts dies inserts flops ushers want carefully heaves
dives chases crumples keys shuffles stuffs went crouches demonstrates strides attaches cranks squints double gingerly tried fishes mounts explains fires tests revs guides taps reveals
won advances felt surveys signals enjoys killed launches made first overhears blasts bangs charges took spreads asked mimes see pumps motions positions smacks simply pockets
parks tips trots wishes roars staggers flexes rounds aims saunters guns lands clamps extracts grits forgets misses paces going rides grins growls screws steers stamps plummets
chuckles loosens explodes dismounts pins circles hammers slashes limps vomits strains yells mumbles chews grunts ponders strolls whistles crashes bounds salutes gropes displays
replaces lost smashes spills loads fails stole barks ambles exchanges bows drains figures itches rears vaults kills howls shot gallops straps

EVEN
MORE "SHE"  MORE "HE"

snuggles giggles squeals sobs weeps blushes clings rocks shrieks hugs shrinks gasps responds trembles pets flinches arches skips utters shudders startles buries swats murmurs resists
hovers caresses awakens shivers screams dances beats absently flees cleans stirs straddles cries moans bites realises mouths accepts wore smiles laughs wrote serves scoots liked
arranges scampers storms twirls softens ignores softly faints wonders fades sags hesitates casts applies hisses fiddles kisses sings awkwardly smokes stretches sips unbuttons stiffens
hurriedly hurries dries look locates threw smooths nods stubs instantly reappears wraps clearly refuses tiptoes lingers beams pivots curls glides strokes meant abruptly retrieves bursts
rakes relents reluctantly frantically understands fidgets breaks recoils hums looked pretends trails frowns retreats gonna licks touches reacts nearly sighs backs embraces squirms panics
yelps ends allows flashes senses exits gathers collects receives drapes folds undoes answers shakes instinctively replies told freezes resumes creeps calms gives rushes sails tentatively
thrashes becomes types avoids clips struggles fight considers leaves thumbs cradles covers blinks dabs meets rests regards tilts attacks darts eyes brushes descends gently nervously
returns forces closes fixes inhales flings scoops desperately plunges appears glimpses clicks halts wakes bolts slaps fumbles quickly wears clasps faces feeds barely shrugs believes
floats left whacks blows immediately emerges seems slips almost stares tugs happily hates slices runs leans sounds washes swallows cranes observes accidentally marches rifles
squeezes begins kneels turns leads snarls rolls wipes whirls tell flushes peeks shuts sits grimaces frees put calls glares tucks like plops scurries whispers tries remains actually
continues pinches tells lets yawns disappears heads looks chokes discovers plugs springs watches cautiously opens clutches studies dropped massages obeys suddenly loves scowls
crosses packs scribbles spits sneaks puts likes just lashes topples hangs lies angles starts claps adds flicks slowly cups angrily really stops pours jabs traces unzips crawls died
grasps slugs steadies breathes glances pushes directs inches hands reads comes unfolds winds attempts rub snorts walks drifts sinks hides goes swivels feels keeps sprays sways
means ducks races repeats used steps sends finishes talking trips locks waits gets snatches chooses obviously takes decides plucks slides moves peers holds claws already stomps
swipes asks admires met said stands buttons owns says sets strips must wanted focuses fell unwraps edges unlocks remembers shines reaches jumps always places climbs stumbles
handles leafs needed passed surreptitiously concentrates helps interrupts reels scrambles steels blocks clenches floors gags splashes rips notices knocks enters finally rises listens
quietly jerks bumps wants lays kicks flies makes picks throws casually scans winces might dangles hefts flails waves dresses realizes passes catches plants thinks knows rings empties
hears sweeps may signs wrenches swims shares started recovers hops props tightens indicates hear dashes got peels will silently probably grows whips finds spins pulls switches
lights assumes ties digs hopes wheels lines performs settles tenses sniffs stabs wanders downs pounds notes slams twists shows weaves bends bounces curses expects hastily pokes
can sees acts scrolls brings arrives needs follows get zips manages proceeds deposits hardly strikes exhales still smells came tears yanks lifts knew shifts presses grabs excuses
straightens hustles speeds recognizes carries pays also falls stays tastes drags never speaks dials turned slows relaxes brought dumps sticks changes know come lowers flips bought
sleeps greet registers succeeds now even withdraws cracks writhes saw nudges scratches hobbles steals pauses collapses offers puffs grinds braces thought expected levels gave hits
drops spots lurches clammers ever eventually snaps writes searches gazes approaches selects eases dips cocks groans lives works winks swerves grips fills rummages loved joins
regains wiggles talks beckons gulps maneuvers zooms stalks seizes vanishes points thrusts hauls leaps hit adjusts heard shoots refers deals honks rams releases clears nears pries
hurls mutters extends sucks trudges found removes accelerates orders produces inspects tumbles cuts calmly presents scrapes crushes pops checks drinks spies drives punches doubles
jams slings smirks fingers slumps prepares wrestles plays counts coughs pats gestures raises tosses lived stoops say contemplates hoists examines rams unscrews uses draws squats
shoves sprints rattles controls hooks surfaces loses delivers lunges dodges addresses called consults eats falters forgot swings shouts dies inserts flops ushers want carefully heaves
dives chases crumples keys shuffles stuffs went crouches demonstrates strides attaches cranks squints double gingerly tried fishes mounts explains fires tests revs guides taps reveals
won advances felt surveys signals enjoys killed launches made first overhears blasts bangs charges took spreads asked mimes see pumps motions positions smacks simply pockets
parks tips trots wishes roars staggers flexes rounds aims saunters guns lands clamps extracts grits forgets misses paces going rides grins grows screws steers stamps plummets
chuckles loosens explodes dismounts pins circles hammers slashes limps vomits strains yells mumbles chews grunts ponders strolls whistles crashes bounds salutes gropes displays
replaces lost smashes spills loads fails stole barks ambles exchanges bows drains figures itches rears vaults kills howls shot gallops straps

Example: Gender Tropes

- Question:
Does our culture, as reflected in movies, promotes certain gender roles?
- Did we answer the question?

Example: Gender Tropes

- Question:
Does our culture, as reflected in movies, promotes certain gender roles?
- Did we answer the question?
 - Is there enough evidence? (Inference, later in the course)

Example: Gender Tropes

- Question:
Does our culture, as reflected in movies, promotes certain gender roles?
- Did we answer the question?
 - Is there enough evidence? (Inference, later in the course)
 - Did we really examine data that answers our question?

Example: Gender Tropes

- Question:
Does our culture, as reflected in movies, promotes certain gender roles?
- Did we answer the question?
 - Is there enough evidence? (Inference, later in the course)
 - Did we really examine data that answers our question?
 - Perhaps it answers:
“Does our culture, as reflected in movies, **reflects** certain gender roles?”

Example: Gender Tropes

- Question:
Does our culture, as reflected in movies, promotes certain gender roles?
- Did we answer the question?
 - Is there enough evidence? (Inference, later in the course)
 - Did we really examine data that answers our question?
 - Perhaps it answers:
“Does our culture, as reflected in movies, **reflects** certain gender roles?”
 - Can you think of different data that would help answer the question?

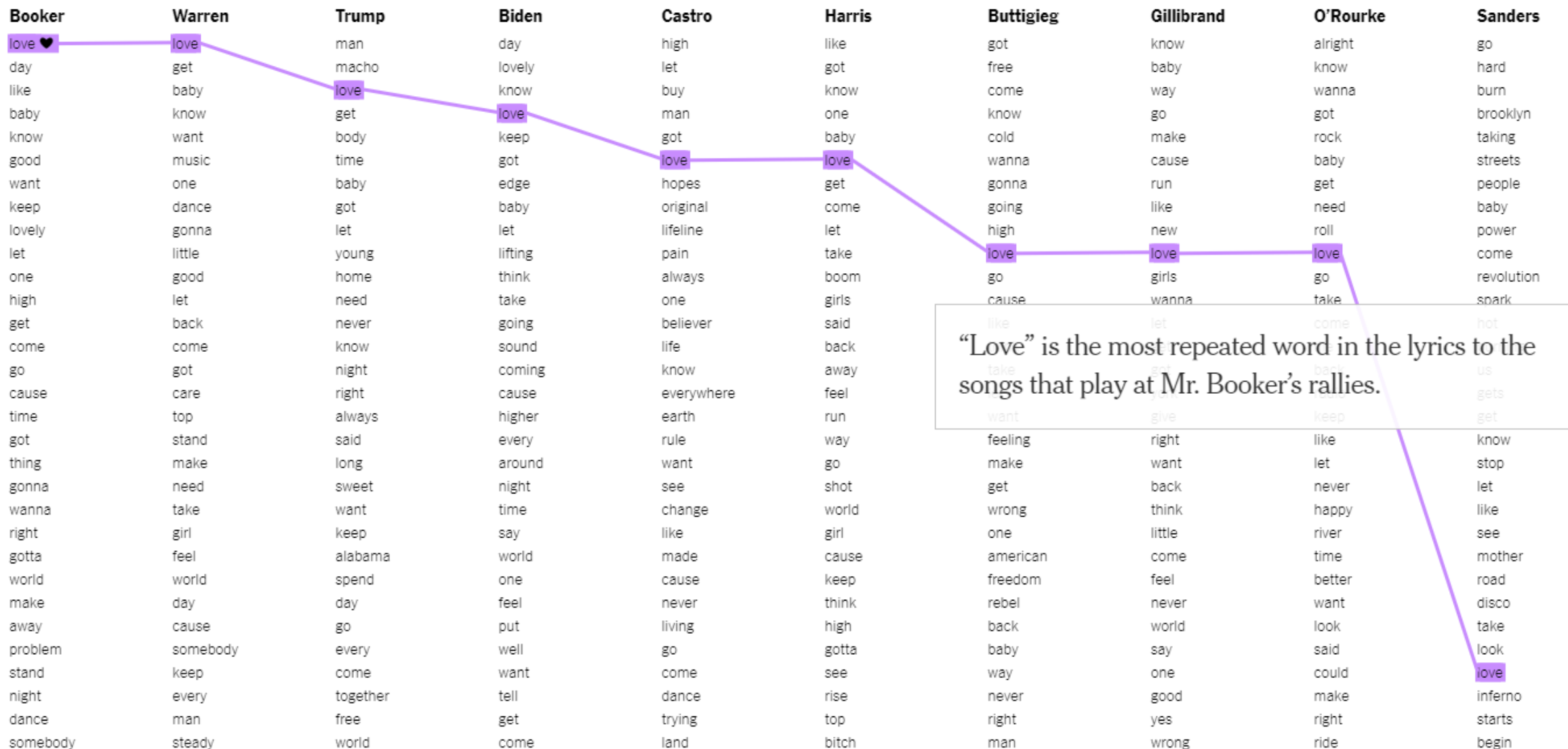
Example: Politician Songs

- Question:
What are the main messages 2020 US presidential race politicians want to associate themselves with?

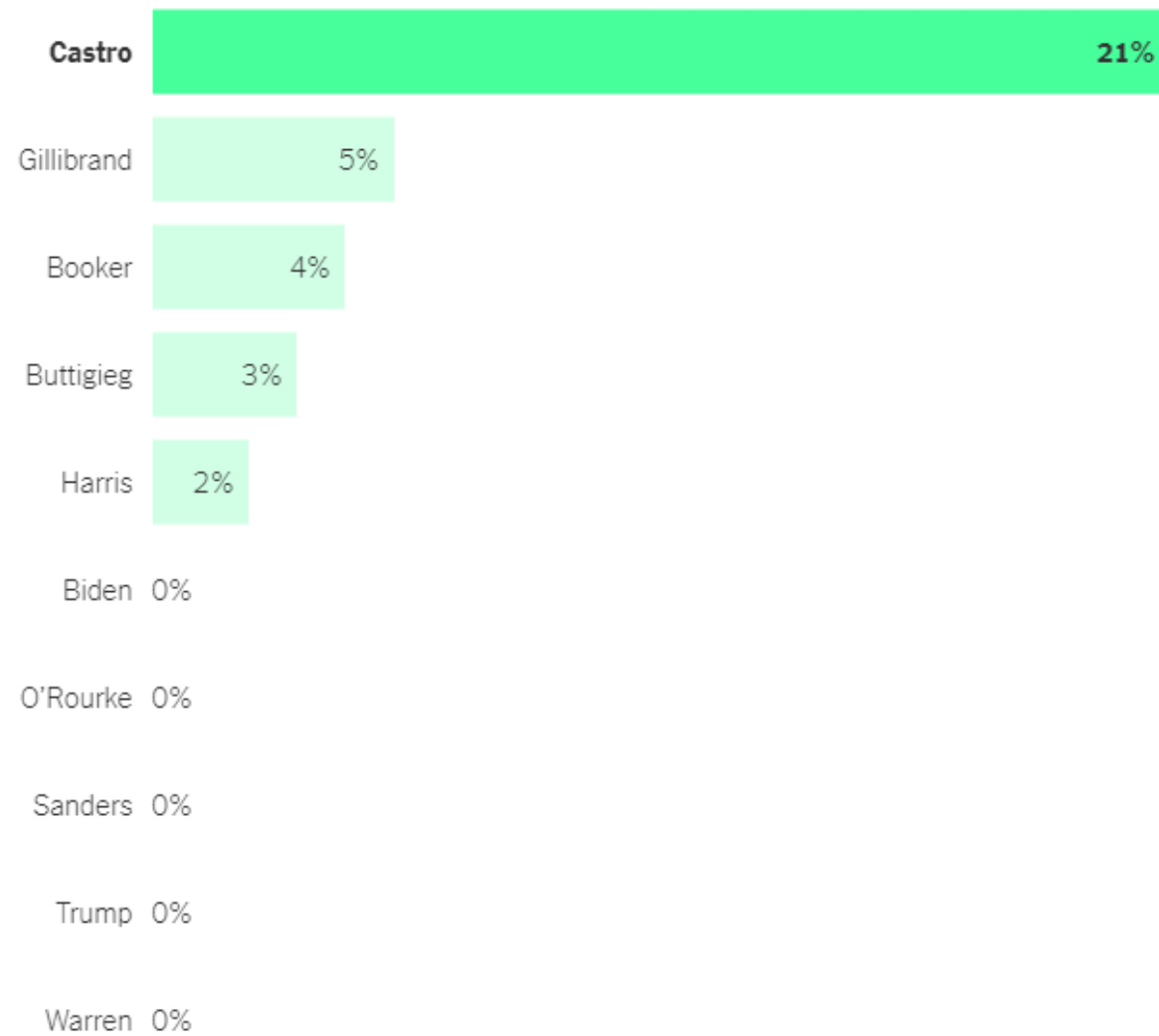
- Question:
What are the main messages 2020 US presidential race politicians want to associate themselves with?
- Data: Songs played in their political rallies



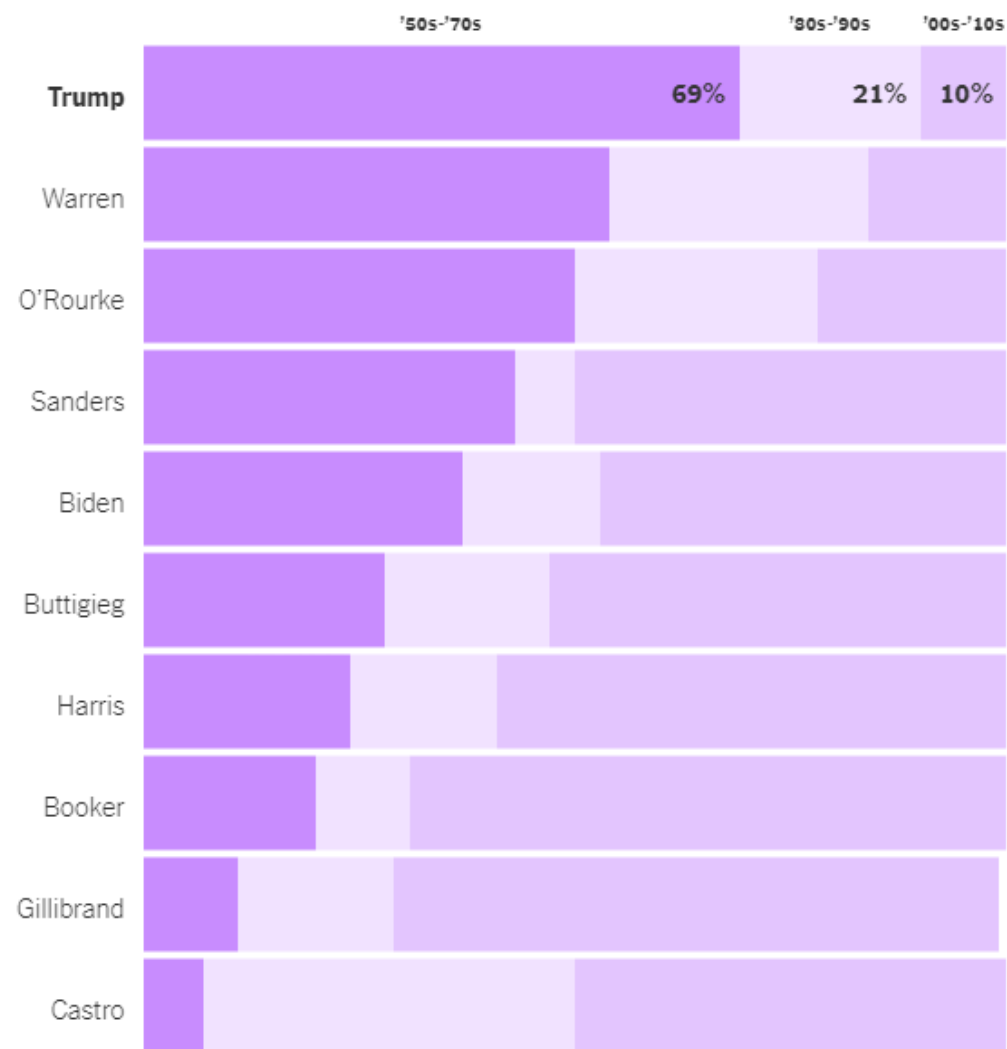
Most frequent words used in each playlist's lyrics



Percentage of Latin songs in each playlist



Percentage of songs by decades



Mr. Trump had the largest share of songs from the previous century, beating Ms. Warren.

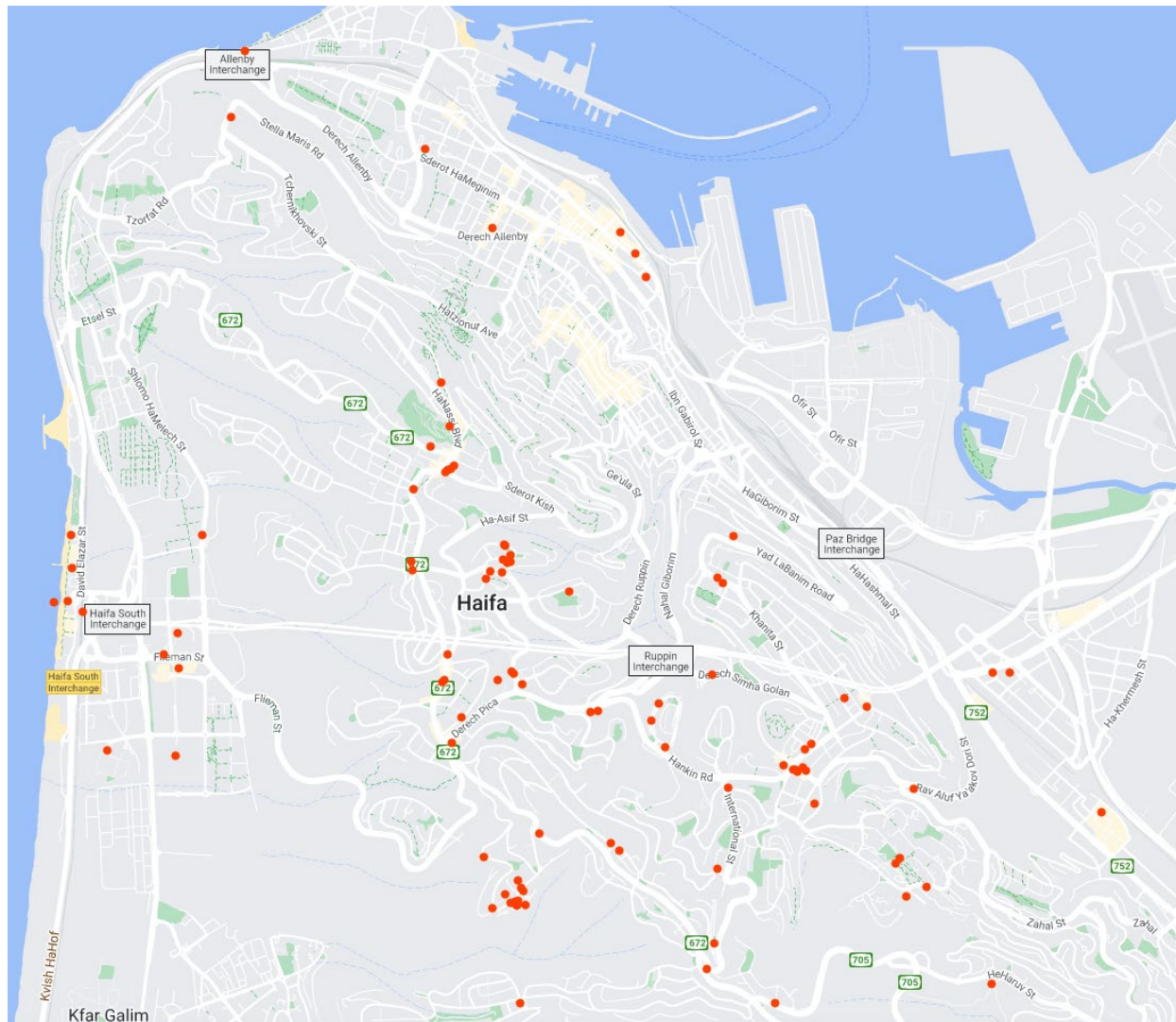
God Bless the USA **LEE GREENWOOD** • Sweet Home Alabama **LYNYRD**
SKYNYRD • Tiny Dancer **ELTON JOHN** • Eye of the Tiger **SURVIVOR** • All Summer
Long **KID ROCK** • Carry on Wayward Son **KANSAS** • We Will Rock
You **QUEEN** • Free Bird **LYNYRD SKYNYRD** • The Best **TINA TURNER** • Wild
Horses **THE ROLLING STONES** • My Way **FRANK SINATRA** • All Right
Now **FREE** • Candle in the Wind **ELTON JOHN** • The Prayer **ANDREA BOCELLI AND**
CÉLINE DION • Ladies Love Country Boys **TRACE ADKINS** • Wheel in the
Sky **JOURNEY** • Turandot / Act 3: "Nessun dorma!" **GIACOMO**
PUCCINI • Y.M.C.A. **VILLAGE PEOPLE** • Under Pressure **QUEEN** • Funeral for a
Friend / Love Lies Bleeding **ELTON JOHN** • Sympathy for the Devil **THE ROLLING**
STONES • Time Is on My Side **THE ROLLING STONES** • Let's Spend the Night
Together **THE ROLLING STONES** • Macho Man **VILLAGE PEOPLE** • We Are the
Champions **QUEEN** • Heart of Stone **THE ROLLING STONES** • Memory **BETTY**
BUCKLEY • Rocking in the Free World **NEIL YOUNG** • You Can't Always Get What
You Want **THE ROLLING STONES**

Mr. Trump's rally music is overwhelmingly performed
by white artists or majority-white bands.

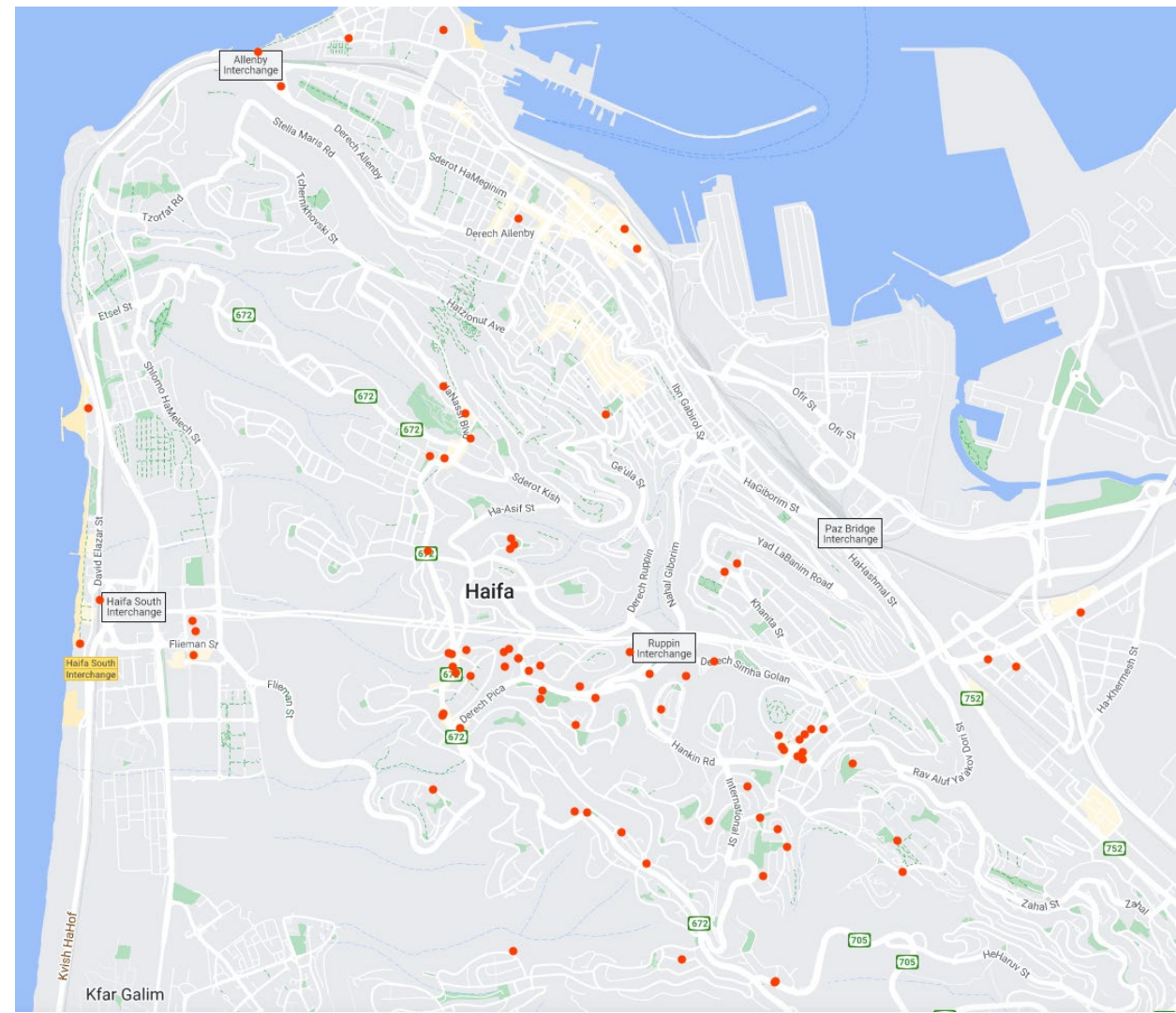
Example: Politician Songs

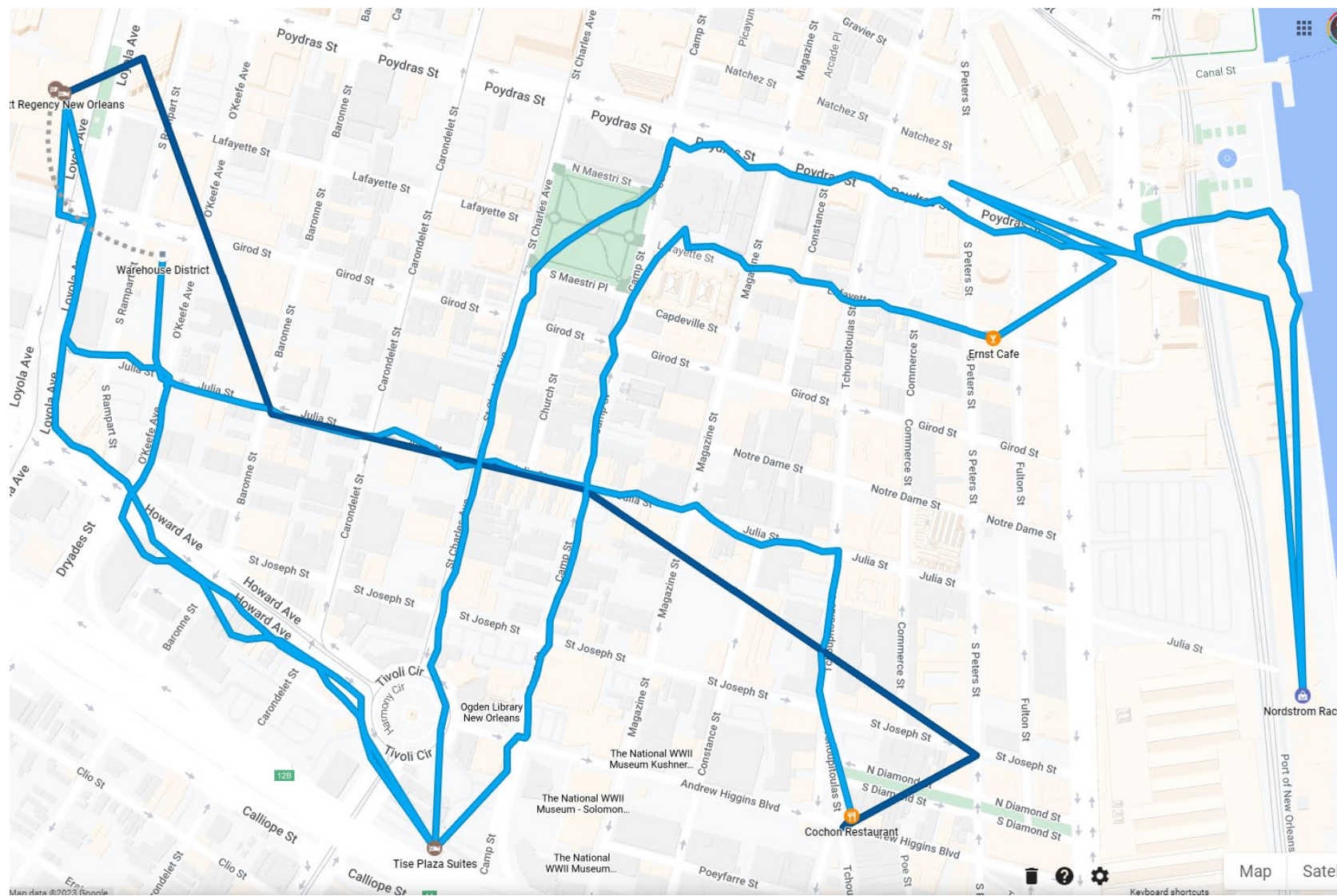
- The NYTimes analyzed the songs for:
 - Genre
 - Race of lead singer
 - Gender of lead singer
 - Word frequency
 - Year
- Can you think of other ways to analyze these data?
 - Do you need more data? Different data?

2018

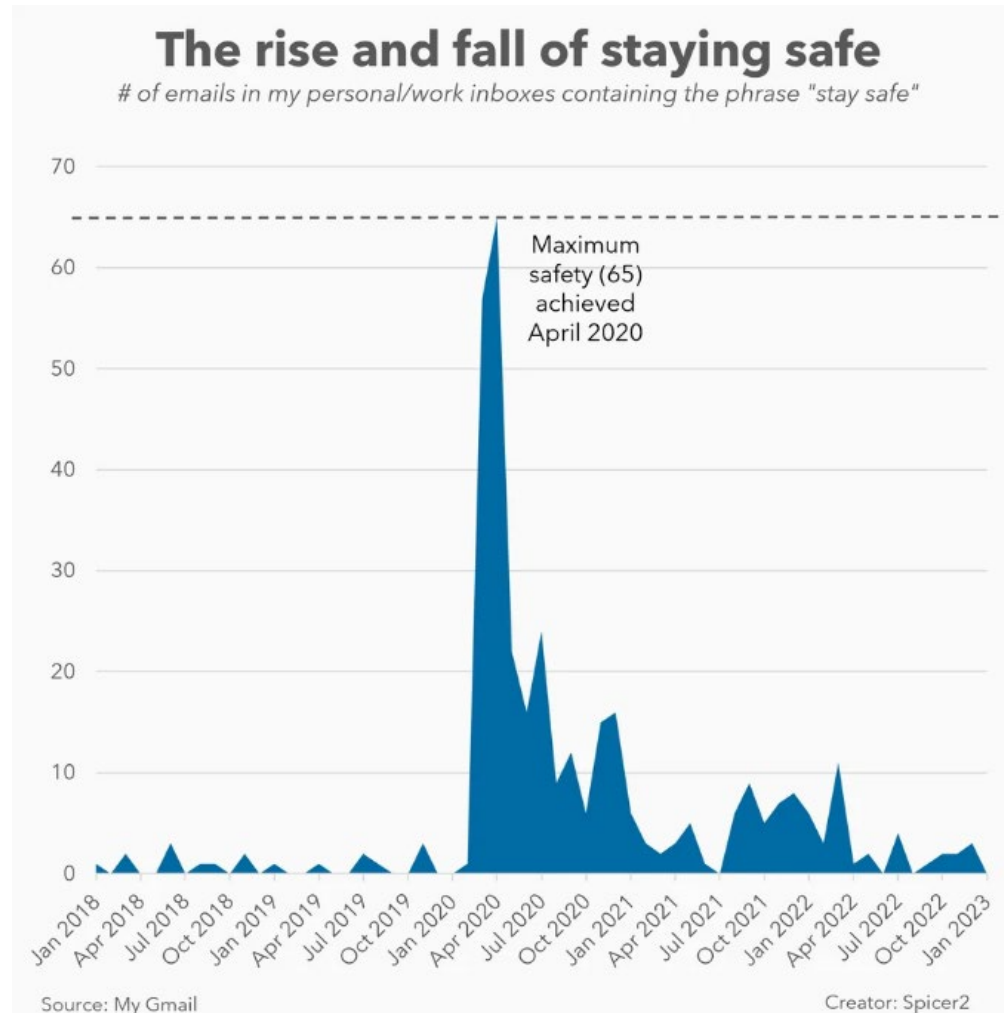


2022





Example: staying safe



Let's play with some data

- Classic Books example
 - Jupyter Notebook

Now, you play with some data

- Run Jupyter Notebooks
 - If you installed Anaconda: Anaconda prompt -> Jupyter notebook
- Open the file *World Happiness exercise.ipynb*
- You will need to have the data file (*happiness 2005_2022.csv*) in the same folder as the *World Happiness exercise.ipynb* file

מדד האושר העולמי: ישראל עלתה בדירוג למקום 4; פינלנד שוב הכי מאושרת

פינלנד היא המדינה המאושרת בעולם, זאת השנה הששית ברציפות, על פי מדד האושר העולמי ל-2023, שכולל נתונים מאנשים ביותר מ-150 מדינות. ישראל עלתה למקום הרביעי השנה, לעומת מקום תשיעי בשנה שעברה

This course

We will learn

- Data processing, “wrangling”
- Exploratory data analysis
- Classification and clustering (simple)
- Pitfalls and subtleties of data analysis
- Effective communication & writing
- Ethical considerations

We will **not** learn (but you will in other classes)

- Hard-core probability and statistics
- Hard-core machine learning

Logistics

Frustrated professor rips off his shirt in the middle of lecture...

🕒 16 October, 2013 in [Pictures](#) | [Comment](#)

In response to a particularly stupid question, my professor ripped off his shirt in the middle of lecture...

