# Adopting computational thinking in data stories

Making the transition
to data journalism

Presented by Hanna
Duggal

ForSet

# Webinar outline

> What is computational thinking

> Self-identity, theory and practice

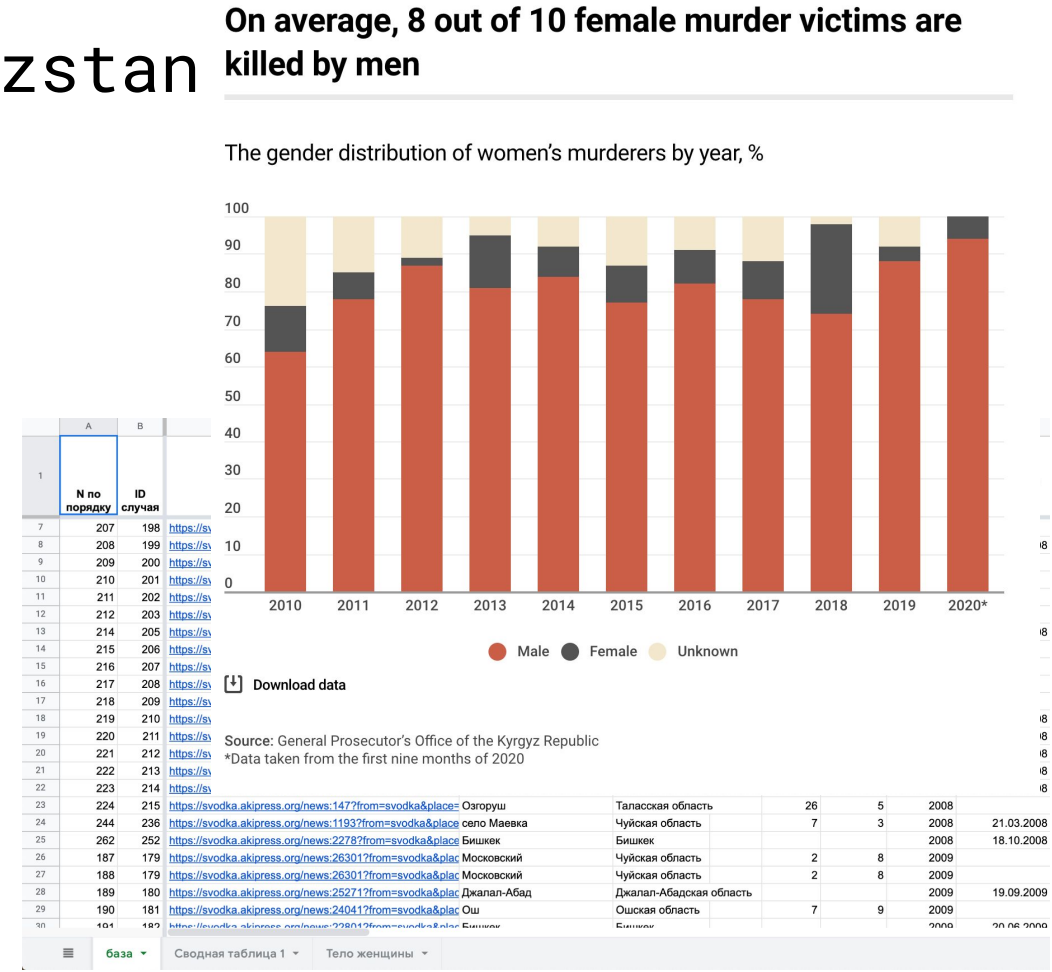
> Paradigms of thinking and toolkits to use

# Femicide in Kyrgyzstan

[“I would have killed her anyway” Kloop's investigation of femicide in Kyrgyzstan]

> Journalists at Kloop looked at the prevalence of femicide in Kyrgyzstan

> How did they go about telling this story?



# Computational thinking

> *“The thought processes involved in formulating problems and their solutions so that the solutions are represented in a form that can be effectively carried out by an information-processing agent”*
>
> **Jeannette Wing**

# Computational thinking

*"The thought processes involved in formulating problems and their solutions so that the solutions are represented in a form that can be effectively carried out by an information-processing agent"*

**Jeannette Wing**

> Computational thinking is analytical

# Computational thinking process

# Decomposition


> Pulling apart steps of a process

> Compositing into smaller steps

> Solvable by computer or human

# Abstraction


> Pulling apart steps of a process

> Compositing into smaller steps

> Solvable by computer or human

> Focusing on what parts need to be solved

> The layer(s) of interest in your data

# Pattern recognition


> Pulling apart steps of a process

> Compositing into smaller steps

> Solvable by computer or human

> Focusing on what parts need to be solved

> The layer(s) of interest in your data

> Recognising the shared characteristics in some elements

> Finding the themes and patterns

# Algorithms


> Pulling apart steps of a process

> Compositing into smaller steps

> Solvable by computer or human

> Focusing on what parts need to be solved

> The layer(s) of interest in your data

> Recognising the shared characteristics in some elements

> Finding the themes and patterns

> How would I get a computer to solve this?

> Writing instructions in the form of functions/scripts

# Where to start?

> The idea

What is the story?

Why does it matter?

> Hypothesis-driven vs data-driven

# Hypothesis-driven vs data-driven

## Hypothesis-driven

> A strong hypothesis

> Confirming or disconfirming a lead

> Sampling

## Data-driven

> A light hypothesis

> Exploring information to find new leads

> Completeness of data

# Story angles

> Scale - identifying a big problem

> Change - things going up/down

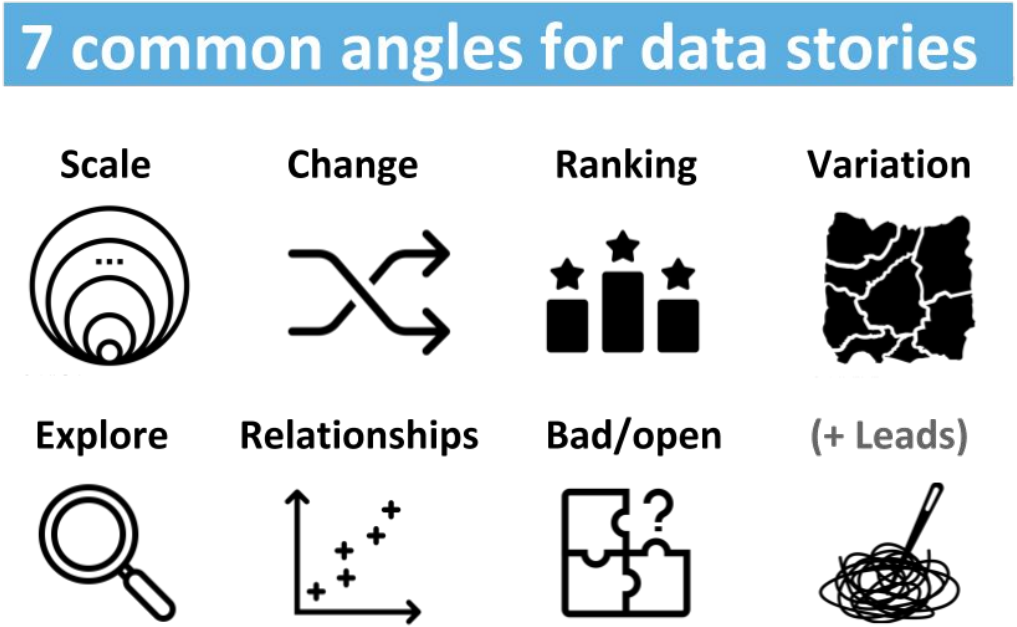
> Ranking - best/worst in dataset

> Variation - expect equal treatment

> Explore - 'call to action'

> Relationships - 'x' related 'y'?

> Bad/open - stories about data itself i.e. lack of data




Credit: Paul Bradshaw

# Example: using police data

> Data-driven approach

> Light hypothesis

> Story angle: Variation + change

## Met police more likely to use force against BAME people

Hanna Duggal Jan 21 · 4 min read

Footage of Met police using force. Source: Channel 4 News

**Metropolitan Police are more likely to use force against BAME people, the latest data reveals.**

# Example: using police data


> FOIs sent to 43 constabularies in England and Wales

> Combining datasets

> Removing fields not of interest in R

> Removing duplicate entries in R

> Sorting the data in R

> Data analysis in Excel

> Comparing 2020 use of force with the previous year, showed use of force against BAME people had gone up.

> Simple data visualisations

> Choice to use a barplot

> Published on Medium

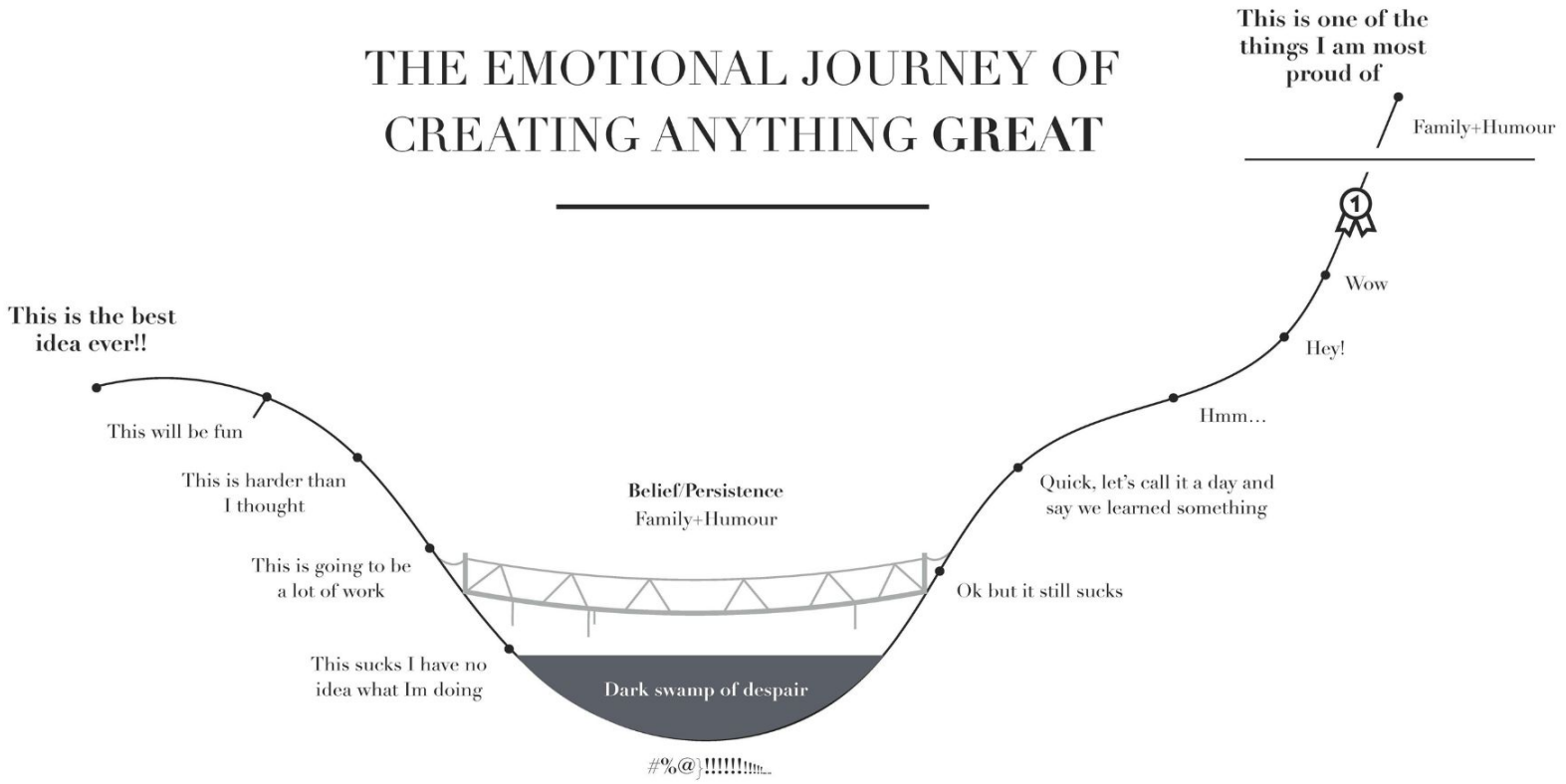
> What would I do differently?

# Self-identity, theory and practice

```
curl 'https://3tzau2bnpu-dsn.algolia.net/1/indexes/*/queries?
20(4.8.0)%3B%20Browser%3B%20instantsearch.js%20(4.2.0)%3B%20J
  -H 'Connection: keep-alive' \
  -H 'x-algolia-application-id: 3TZAU2BNPU' \
  -H 'x-algolia-api-key: f3f54e2629fb29af51828f984180265b' \
  -H 'User-Agent: Mozilla/5.0 (iPhone; CPU iPhone OS 13_2_3 l
  like Gecko) Version/13.0.3 Mobile/15E148 Safari/604.1' \
  -H 'content-type: application/x-www-form-urlencoded' \
  -H 'Accept: */*' \
  -H 'Origin: https://www.elysee.fr' \
  -H 'Sec-Fetch-Site: cross-site' \
  -H 'Sec-Fetch-Mode: cors' \
  -H 'Sec-Fetch-Dest: empty' \
  -H 'Referer: https://www.elysee.fr/' \
  -H 'Accept-Language: en-US,en;q=0.9,he;q=0.8' \
  --data-raw '{"requests":[{"indexName":"prod_all","params":"
  T%20type%3Aevent&highlightPreTag=__ais-highlight__&highligh
  uesPerFacet=20&facets=%5B%22categories.fr%22%2C%22president
  22categories.fr%22%2C%22topics.fr%22%2C%22president.name%22
  es.fr%3ADiscours%22%2C%22president.name%3ANicolas%20Sarkozy
  --compressed
```

“What the meow is this?”

# THE EMOTIONAL JOURNEY OF CREATING ANYTHING **GREAT**




THE EMOTIONAL JOURNEY IS **INEVITABLE** AND PERHAPS NECESSARY

# Passing the threshold

> Transformative - a significant change in perception

> Irreversible - unlikely to be forgotten

> Integrative - exposing interrelatedness of something

> Troublesome - troublesome knowledge

Meyer and Land, 2005

# Paradigms of thinking

> How we view the world

> Thought patterns, assumptions, methods, contributions

1. How to find or access data
2. Data literacy: descriptive and predictive statistics, significance, etc.
3. Coding of big data
4. Data visualisation: telling stories using charts, maps, infographics

"I know how to break down a problem into smaller problems when it comes to something technical"

"Perform operations on a dataset manually and intuively. Then generalise your intuition into a systmatic approach using programming."

"I usually start sketching with words first, very light initial ideas for outputs and then shift towards mapping the data experience journey in a modular way that I can disassemble and assemble"

"Your dataset is kind of like a source. Your job is to interview it like any other source. You understand as much as you can and back it up with other sources"

"You start with a puzzle, a question. It's a permanent process of zooming in and zooming out. Without me saying anything the instinct was to zoom in to certain parts of the network and interrogate those nodes"

"I know how to break down a problem into smaller problems when it comes to something technical"

Web developer

"Perform operations on a dataset manually and intuively. Then generalise your intuition into a systmatic approach using programming."

Programmer

"I usually start sketching with words first, very light initial ideas for outputs and then shift towards mapping the data experience journey in a modular way that I can disassemble and assemble"

Information designer

"Your dataset is kind of like a source. Your job is to interview it like any other source. You understand as much as you can and back it up with other sources"

Journalist

"You start with a puzzle, a question. It's a permanent process of zooming in and zooming out. Without me saying anything the instinct was to zoom in to certain parts of the network and interrogate those nodes"

Social scientist

# Example 1: Sport Climbing

Climbing at the Olympics

> Data-driven approach

> Scraped data from the IFSC

> Visual storytelling


Credit: Niels de Hoog

# Example 1: Sport Climbing

Climbing at the Olympics

> Data-driven approach

> Scraped data from the IFSC

> Visual storytelling

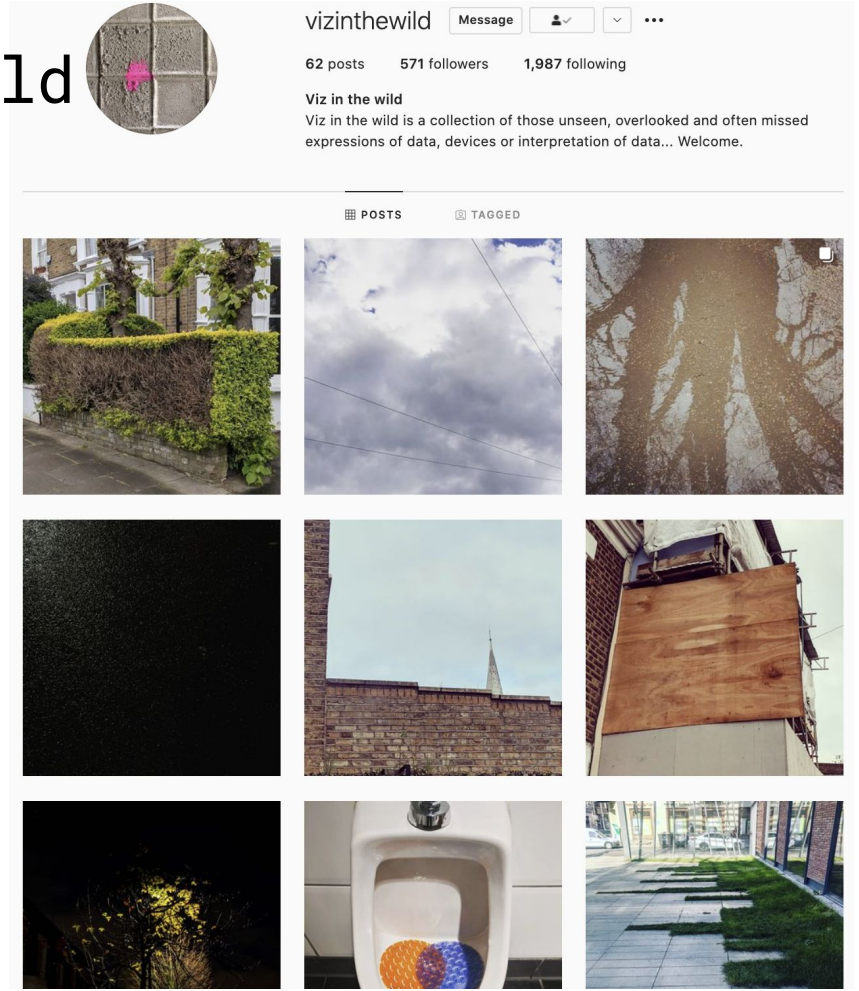
"Try to find something you're passionate about - that will carry you through the difficulty of what you're learning"


Credit: Niels de Hoog

Credit: Martin Brenner

# Example 2: Viz in the wild

>Viz in the wild (@vizinthewild) • Instagram photos and videos




"Viz in the wild was a way for me to visually explore how visualisation types could function and how I might be able to take a visualisation and break it free from flat land only existing in a 2D environment"

# Example 3: Collective Action Networks in Egypt

> Research design

> Convergent parallel approach

> Interviews and creation of dataset

> Question > boundary specification > data collection > data analysis > interpretation (loop)

# Toolkits to use


> Where and how are you obtaining the data?

> Available, requested, assembled, scraped, leaked

> What tools are you using in the cleaning and organising?

> Excel, OpenRefine, R, Python, SQL, Command Line etc.

> How are you understanding the numbers?

> Statistical analyses, validity etc.

> How can you best communicate the data?

> Charts, maps, visuals, scrollytelling, audio etc.

> Has the message been delivered?

> What would you change

> Outputs and conversations

# Collaboration

> What is the value of collaboration?

> Mixed approaches, where and how is value added along the chain?

# Thank you for listening!

Email: hanna.duggal@gmail.com