ANTIPATTERN ASSUMPTIONS IN DATA

Rebecca Long | @amaya30
Future Ada | @futureada
SQL Saturday Spokane 2019 | #SQLSatSpokane

Hello!

- Rebecca Long
- Spokane, WA
- President & Founder of Future Ada
- Lead DevOps Engineer at RiskLens
- Eastern Washington University Double Alumn in Computer Science

Today's Plan

- ■ I'm not going to...
  - – *Provide you with all the answers*
  - – *Tell you how to build your systems*
- ■ I am going to...
  - – *Push you to think outside the box*
  - – *Have you look at data differently*

You have to let it all go, Neo. Free your mind.

What on Earth is an "antipattern"?

- Common solutions to common problems where the solution is ineffective and may result in undesired consequences

- Different from bad practice when:
  - *It is a common practice that initially looks like an appropriate solution by ends up having bad consequences that outweigh any benefits*
  - *There's another solution that is known, repeatable, and effective.*

- Concept inspired by design patterns
  - *Which indicate common effective solutions to common problems.*

Definition from The Agile Alliance

What is a data "antipattern"?

- We spend a lot of time mapping and storing data

- How we store data is based on our perceptions on how data should be categorized

- A "data antipattern" is a flaw in our default categorization of data
  - *Misconceptions in data*
  - *Conscious and unconscious bias*
  - *Cultural preconceptions*

Unconscious Bias

- "Your background, personal experiences, societal stereotypes and cultural context can have an impact on your decisions and actions without you realizing.

  Implicit or unconscious bias happens by our brains making incredibly quick judgments and assessments of people and situations without us realizing. Our biases are influenced by our **background**, **cultural environment** and **personal experiences**. We may not even be aware of these views and opinions, or be aware of their full impact and implications."

https://www.ecu.ac.uk/guidance-resources/employment-and-careers/staff-recruitment/unconscious-bias/
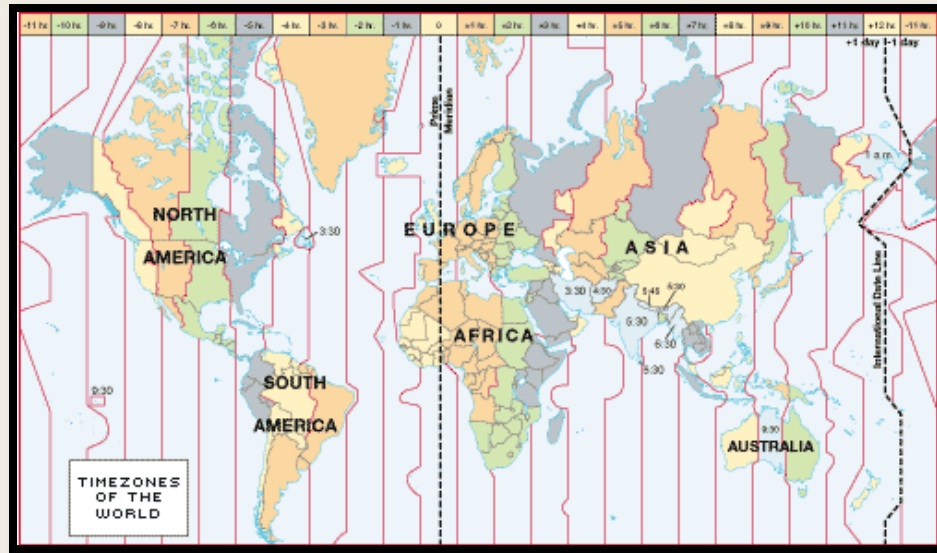
TIME

Time Antipatterns

- The time zone in which a program has to run will never change

- The system clock will never be set to a time that is in the distant past or the far future

More Time Antipatterns

TIMEZONES OF THE WORLD

- A time stamp of sufficient precision can safely be considered unique

- The duration of one minute on the system clock would never be more than an hour

- The local time offset (from UTC) will not change during office hours

- My software is only used internally/locally, so I don't have to worry about timezones

- I can easily maintain a timezone list myself

- One minute on the system clock has exactly the same duration as one minute on any other clock

LOCATIONS

Address Antipatterns

- No buildings are numbered zero

- A road will have a name

- A single postcode will be larger than a single building

- There won't be multiple postcodes per building

- Addresses will have a reasonable number of characters — less than 100

Geography Antipatterns

- Places have only one official name
- Place names follow the character rules of the language
- Place names can be written with the exhaustive character set of a country
- Places have only one official address
- Street addresses contain street names

Map Antipatterns

- All coordinates are in "Latitude/Longitude"
- The shortest path between two points is a straight line
- All programmers agree on the ordering of latitude and longitude pairs
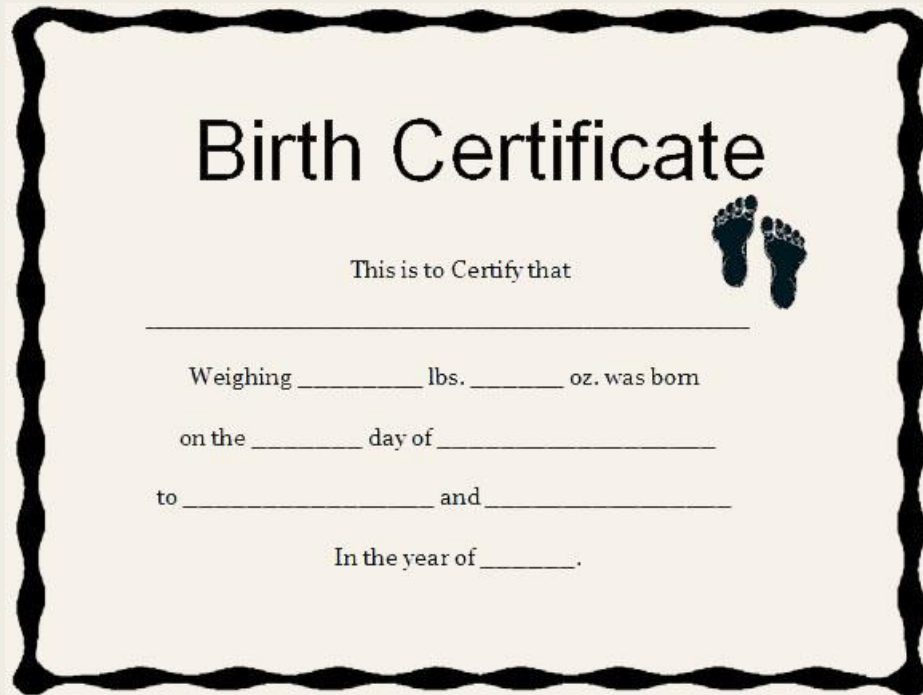
NAMES

Name Antipatterns

- People have exactly one canonical full name
- People use an initial for their middle name
- People have one first, one middle, and one last name to make up their full name
- People have exactly N names, for any value of N
- People's names have an order to them
- People's names fit within a certain defined amount of space
- People's names do not change

- A dictionary of bad words will not contain any names

Split Name Fields vs. Single Name Field

First Name

Last Name   Full Name

✕ Culturally Exclusive   ✓ Culturally Inclusive

More Name Antipatterns

Birth Certificate

This is to Certify that

_____

Weighing _____ lbs. _____ oz. was born

on the _____ day of _____

to _____ and _____

In the year of _____.

- People's first names and last names are, by necessity, different
- People have last names, family names, or anything else which is shared by folks recognized as their relatives
- People's names are globally unique
- People's names are assigned at birth

Even More Name Antipatterns

- Two different systems containing data about the same person will use the same name for that person
- Two different data entry operators, given a person's name, will by necessity enter bitwise equivalent strings on any single system, if the system is well-designed
- People whose names break my system are weird outliers
  - They should have had solid, acceptable names, like 村山裕子 *(Yuko Murayama)*

Antipatterns about the Characters in Names

- People's names are written in ASCII.
- People's names are written in any single character set.
- People's names are all mapped in Unicode code points.
- People's names are case sensitive.
- People's names are case insensitive.

More Character Antipatterns

- People's names sometimes have prefixes or suffixes, but you can safely ignore those
- People's names do not contain numbers
- People's names are not written in ALL CAPS
- People's names are not written in all lower case letters
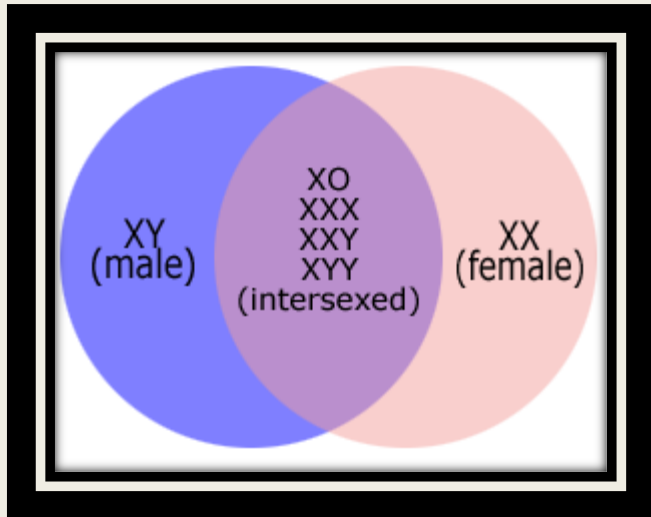
GENDER

Quick Definitions

Sex

- Biological sex
  - *Anatomy of an individual's reproductive system*

Gender

- Gender identity
  - *Personal identification of one's own gender based on internal awareness*

- Gender role
  - *Social roles based on sex of the individual*

https://en.wikipedia.org/wiki/Sex_and_gender_distinction

Biological Sex Antipatterns

- A person has only a single biological sex
- There are only two biological sexes
- Biological sex is clearly and unambiguously defined
- DNA clearly distinguishes males and females

Gender Antipatterns

- People are either male or female
- People will be fine with male being the default option
- People who are not male or female will be happy to be lumped together under "other"
- A person's gender never changes
- Gender is unambiguously defined
- People have a single gender at any given time that they use for all purposes

Gender Antipatterns

- A person's gender is obvious by their appearance or tone of voice

- Is this photo of a male or female?

More Gender Antipatterns

- A person's gender is public information
- A person's gender signifies how they wish to be addressed
- A person's gender signifies what grammatical gender and pronouns to use with them
- A person has one *legal* gender that is consistent across all their forms of identification
- All *legal* forms of identification have a person's gender
- A person's *legal* gender can only be male or female

Gender & Families

- A person has exactly two parents
  - *a mother & a father*
- A person has only two *biological* parent
  - *a mother & a father*

FAMILIES

Family Size

Use Google Play Family Library

You can share purchased apps, games, movies, TV shows, and books from Google Play with up to 5 family members using Google Play Family Library.

https://families.google.com/families

- Approximately 9.5% of US families in 2018 had 5 or more family members
- Approximately 10% of babies born in the UK have 3+ siblings
- What if you wanted to include grandparents or in-laws?

Changes in Family

> **Can't create family**
>
> You can only change family groups once every 12 months. You'll need to wait 12 months until you can create or join a new family.

- What about newly adopted kids who have been in foster families this past year?

- Split from your partner and have to wait a whole year to join your new partner's family?

Family Adults

If you are not part of a family already, you'll become the family manager when you sign up for Family Library. Family managers must meet all these requirements:

- Be 18 or older (20 in Japan).
- Have a valid credit card to use as the family payment method.

- Average legal age of marriage for boys is 17 and 16 for girls
- Many countries allow marriage at even younger ages, particularly girls
- Some US states (e.g. Massachusetts) allow girls as young as 12 to get married in "exceptional circumstances" with consent from a judge

Family Cohabitation

Family member requirements

To join a family group, you must meet these requirements:

- Have a Google Account. If you're under 13, your family manager needs to create a Google Account ☑ for you.
- Live in the same country as the family manager.
- Not be part of another family group on Google.

- Adult kids deployed overseas
  - *Not in the same country as family manager*

- Kids who live part time with Mom & Step-Dad and part time with Dad
  - *do they have to pick only one family to be part of?*

WRAP UP

Now what?

- Do we need to build our applications to support **100%** of *all possible cases*?
  - *No*

What should we do?

THERE IS NO SPOON.

- Be mindful of the data we are collecting, saving, and mapping
  - *What assumptions are being made?*
  - *Is this the data we need?*
  - *Are we excluding people unnecessarily from our system?*
  - *Will we need to change it in the near future?*
  - *Far future?*
  - *Are we okay with this?*

- Be mindful of your own biases that might be coloring how we view the application data we work with
  - *Break down your own stereotypes on data*

Where did all this come from?

- Falsehoods Programmers Believe
  https://spaceninja.com/2015/12/07/falsehoods-programmers-believe/
- Personal Names Around the World
  https://www.w3.org/International/questions/qa-personal-names
- Gay Marriage: The Database Engineering Perspective
  https://qntm.org/gay
- Falsehoods Programmers Believe About Families
  https://shkspr.mobi/blog/2017/03/falsehoods-programmers-believe-about-families/
- Falsehoods Programmers Believe About Names
  https://www.kalzumeus.com/2010/06/17/falsehoods-programmers-believe-about-names/
- Falsehoods Programmers Believe About Gender
  https://medium.com/gender-2-0/falsehoods-programmers-believe-about-gender-f9a3512b4c9c
- Baby Name Rules by State
  https://www.thebump.com/a/baby-name-rules

Thank you! Questions?

Contact Me

- rebecca@futureada.org
- Social:
  - *@amaya30*
  - *@futureada*
- Web:
  - *https://futureada.org*
  - *http://rebeccalong.tech*