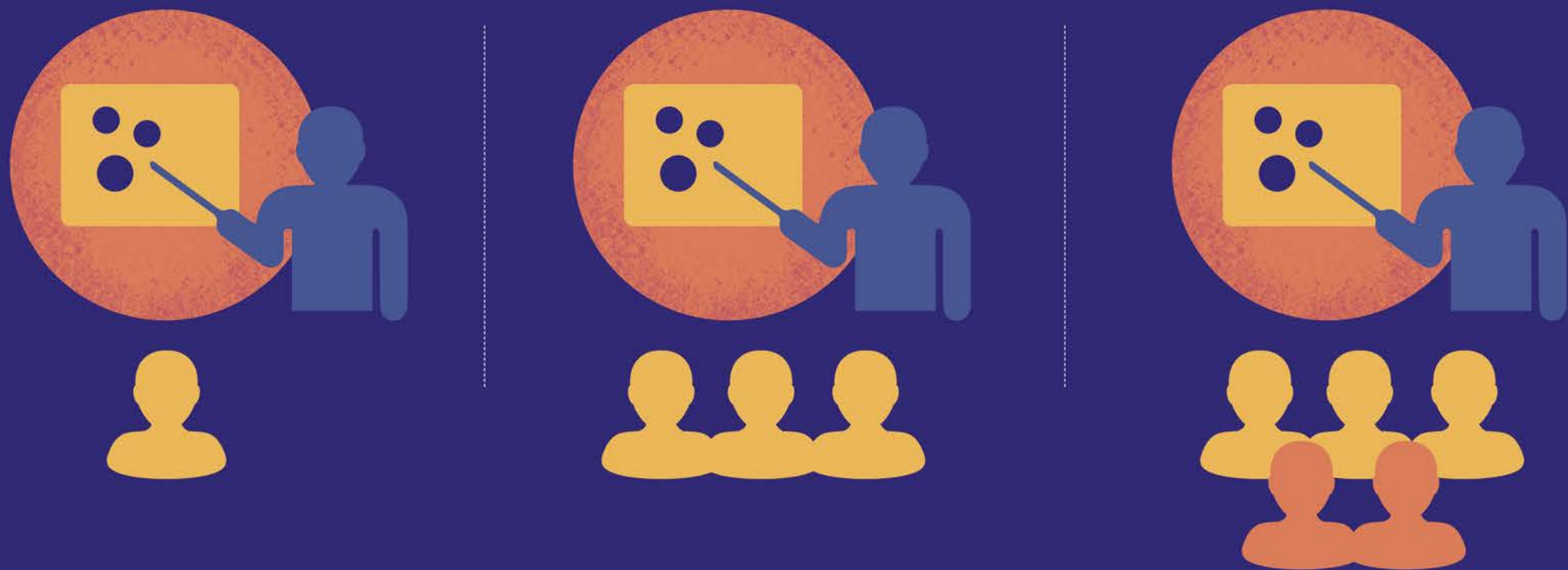
Data Ethics in R

Step by step processes to avoid racism, sexism, homophobia and more in data and analysis.

Heather Krause, PStat

What is the average size classroom?



The average classroom is 3 students per class.

The average classroom is 4 students per class.

Both are correct.

Teacher Perspective

1 + 3 + 5 = 9

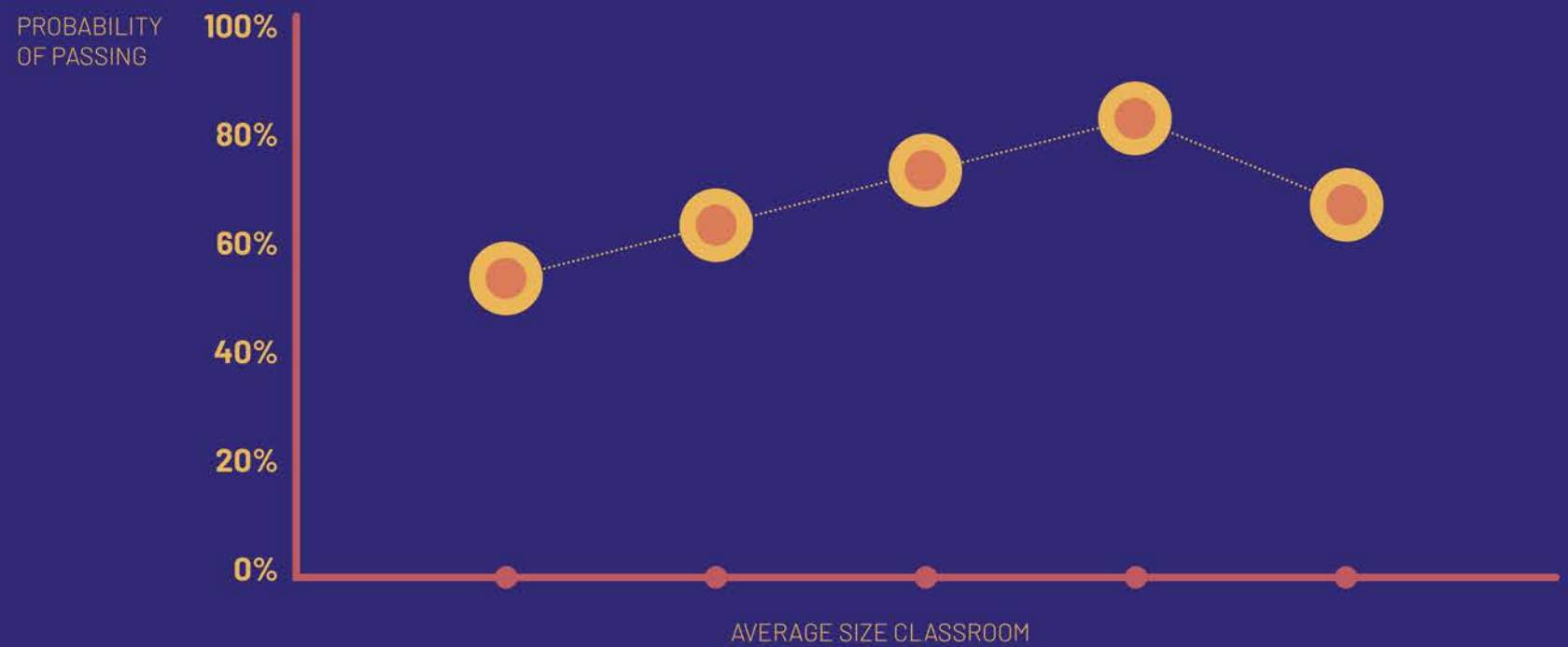
9/3 = 3

Student Perspective

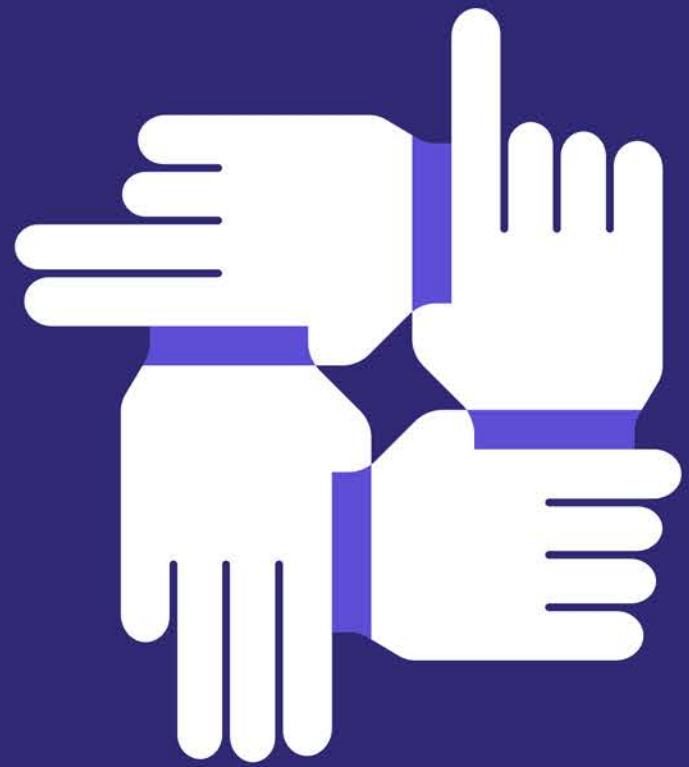
1 + 3 + 3 + 3 + 5 + 5 + 5 + 5 + 5 = 35

35/9 = 4

Here is the relationship between class size and academic performance from the student's POV.



Here is the relationship between class size and academic performance from the teacher's POV.



WE ALL COUNT

project for equity
in data science

What is Bias?

"Systematic error introduced into sampling or testing by selecting or encouraging one outcome or answer over others."

Sources of bias can be identified in each step of the data life cycle.

Funding

Motivation

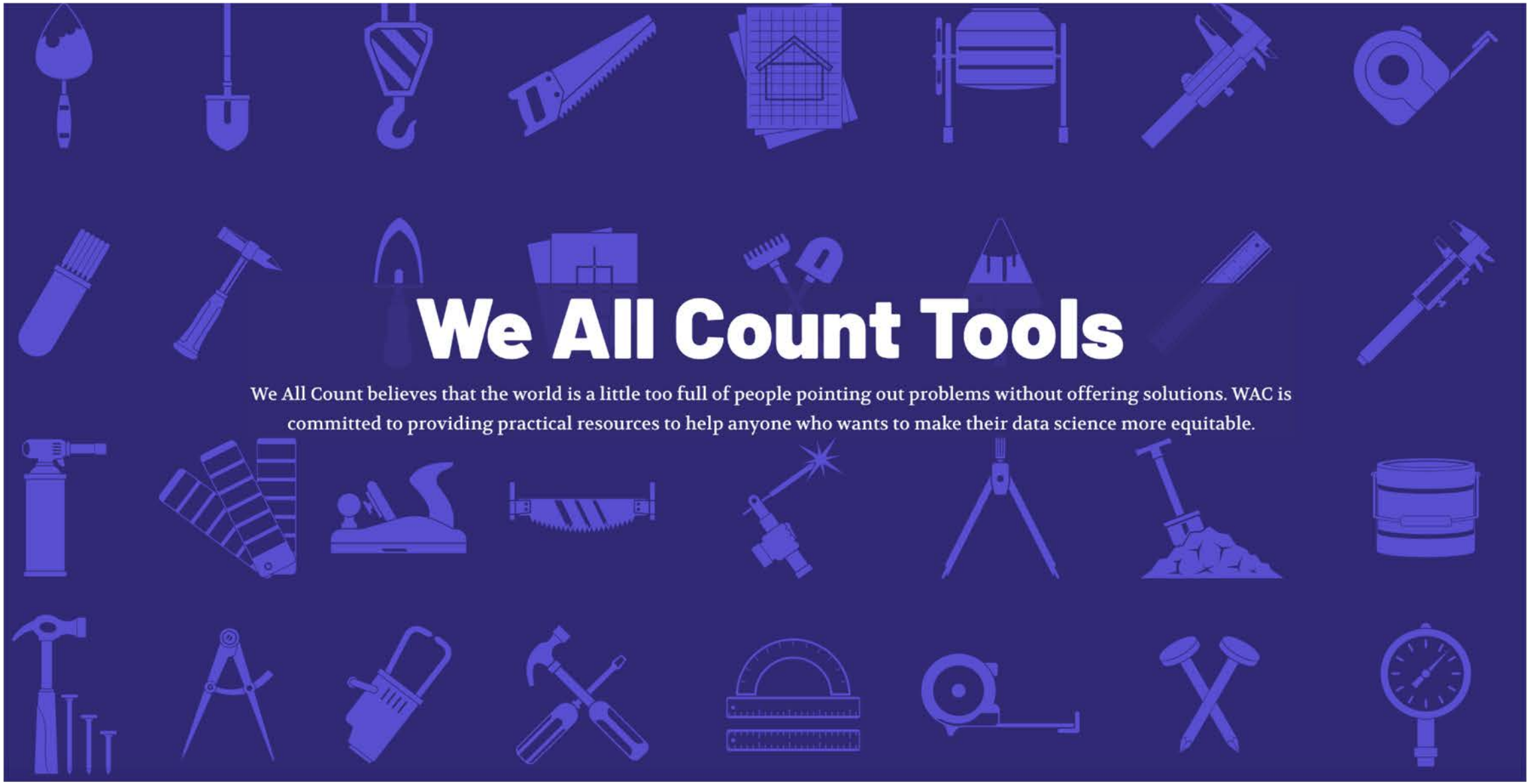
Project Design

Data Collection & Sourcing

Analysis

Interpretation

Comunication & Distribution



We All Count Tools

We All Count believes that the world is a little too full of people pointing out problems without offering solutions. WAC is committed to providing practical resources to help anyone who wants to make their data science more equitable.

Data Ethics and Equity Checklist for Data Science in R. Part of the We All Count project.

Heather Krause, Georges Monette

July 3, 2019

Funding

Step 1: Funding Statement

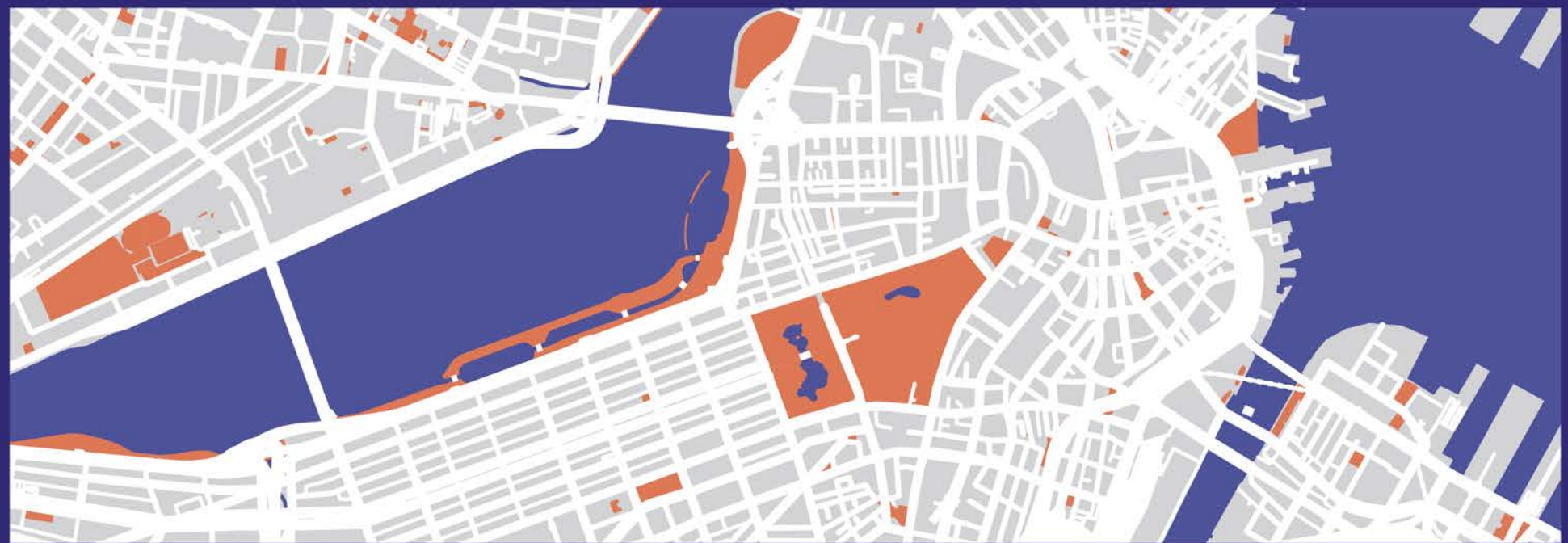
The first step in getting your analysis on strong ethical footing is to include a brief statement about who and what organizations are funding the project. Include as this funding path as far up the chain as possible.

Motivation

Why is this data project being done?

Tension between explicit purpose (understand if this works) and implicit purpose (have a good report for the annual general meeting).

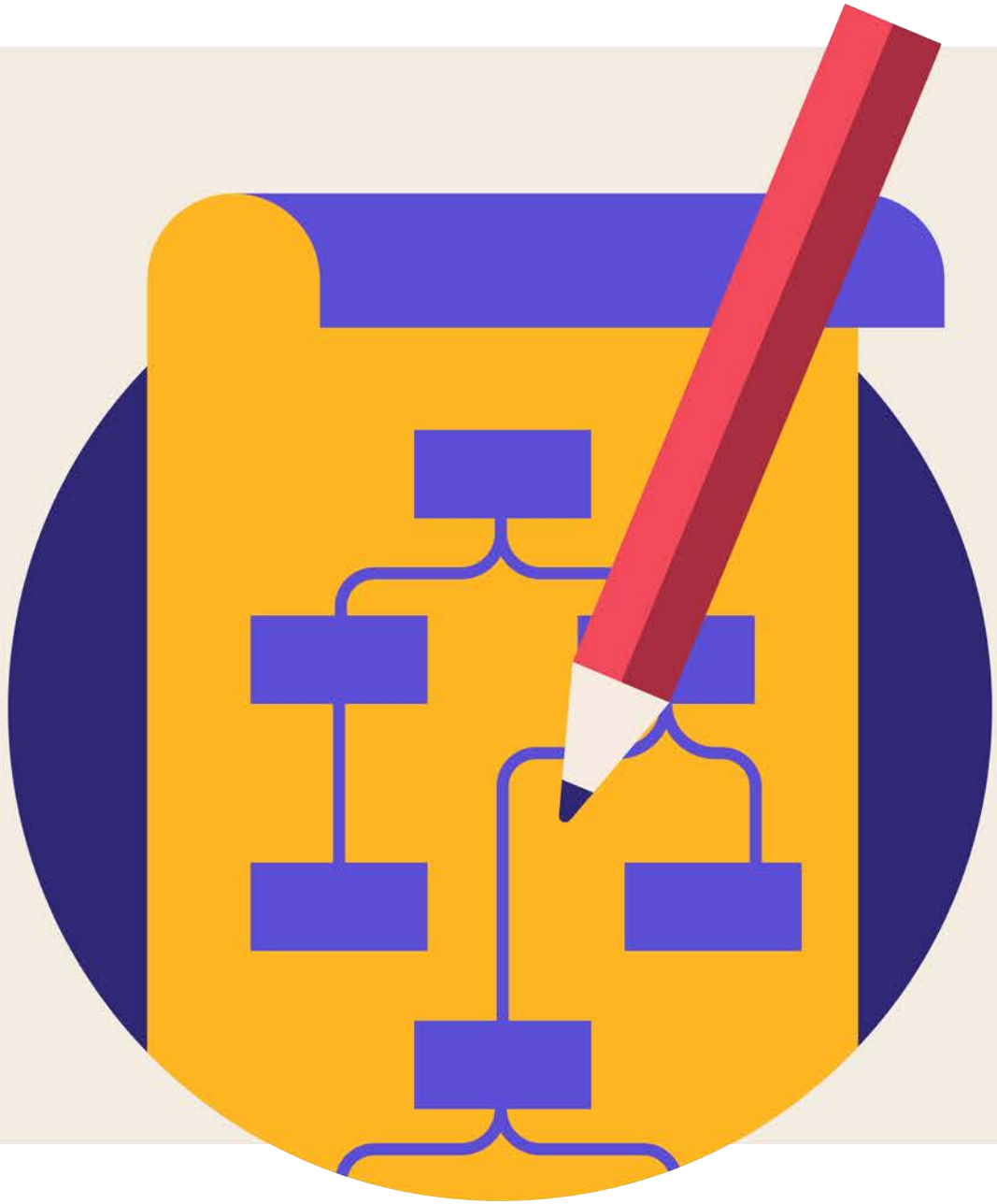
Happy Maps - WHY are you designing this data project?



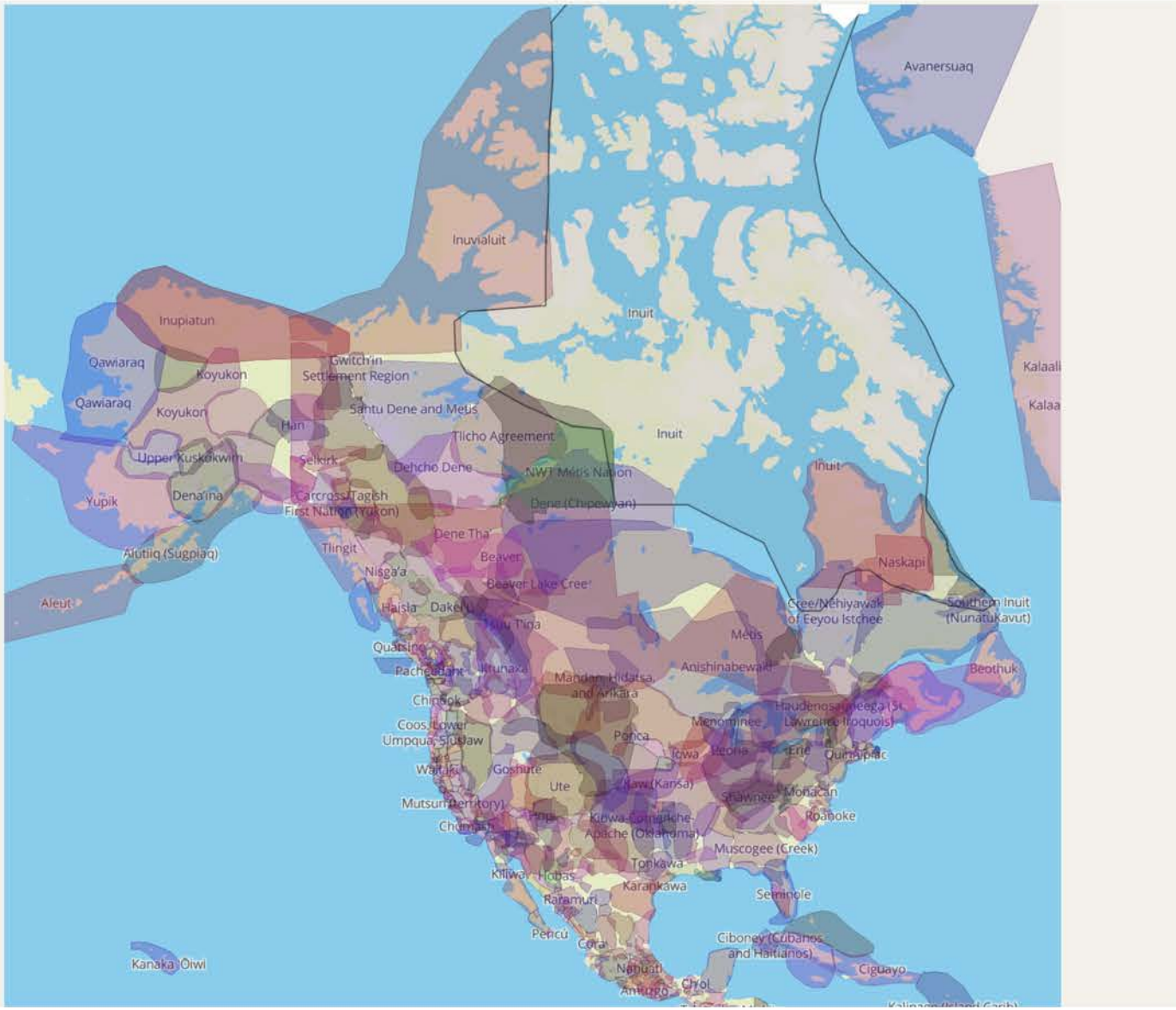
Step 2: Motivation Statement

The next step is to include the data product's Motivation Statement. This clearly lays out both the explicit and implicit motives and hopes and dreams of your data analysis. It's a crucial step in laying the foundation of an ethically strong piece of work. If you'd like some guidance you can download the Motivation Statement checklist

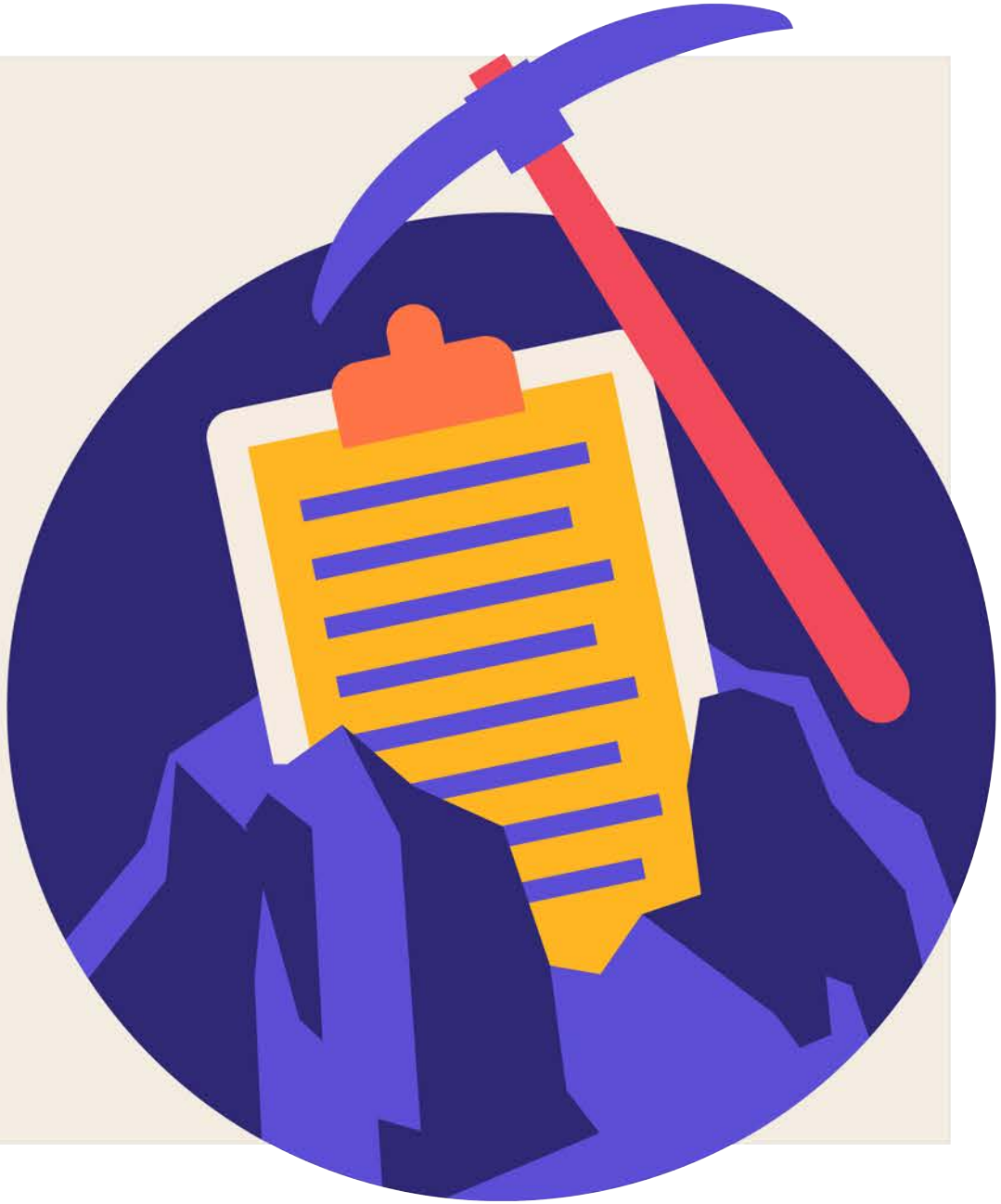
Project Design

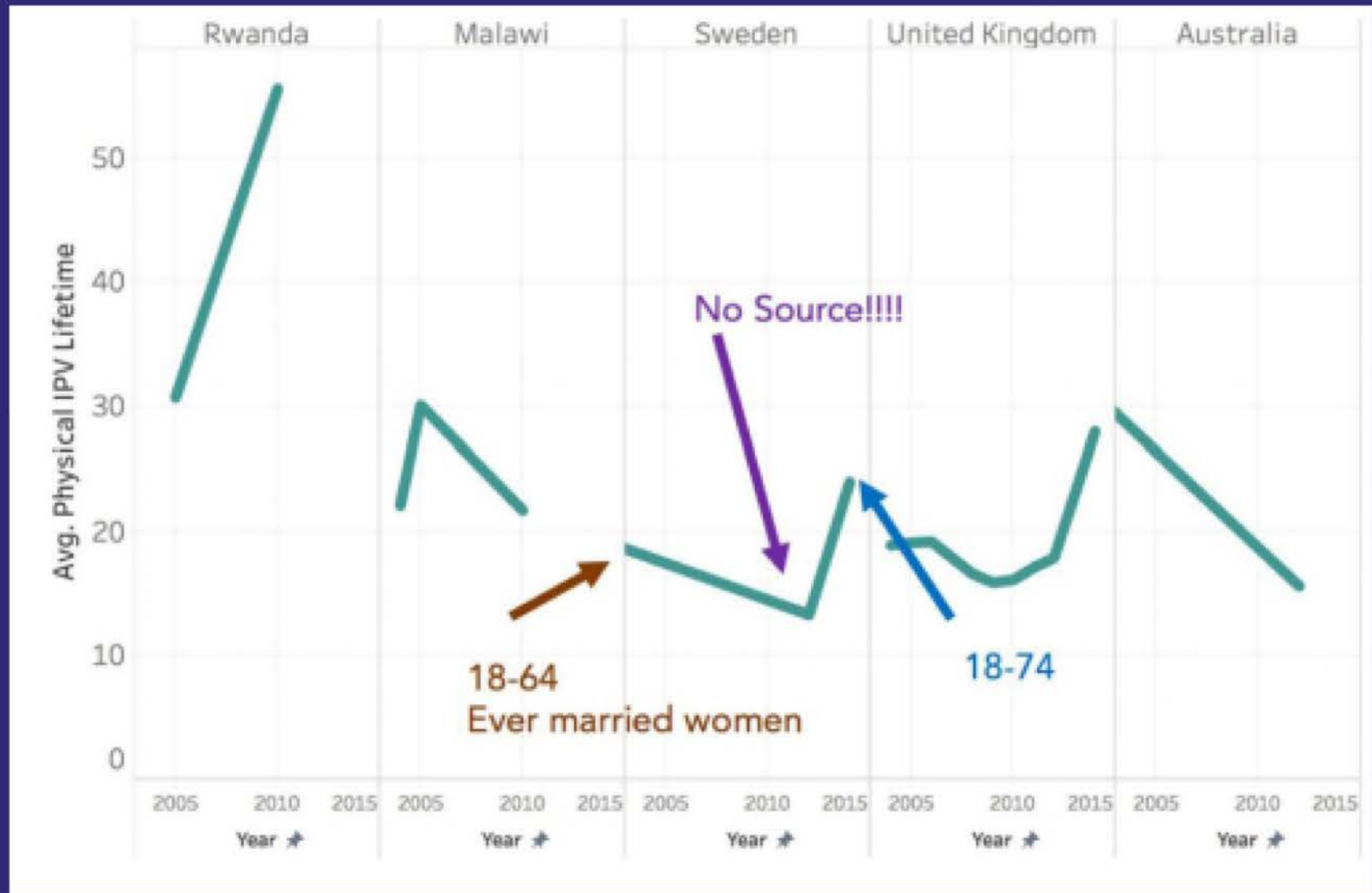


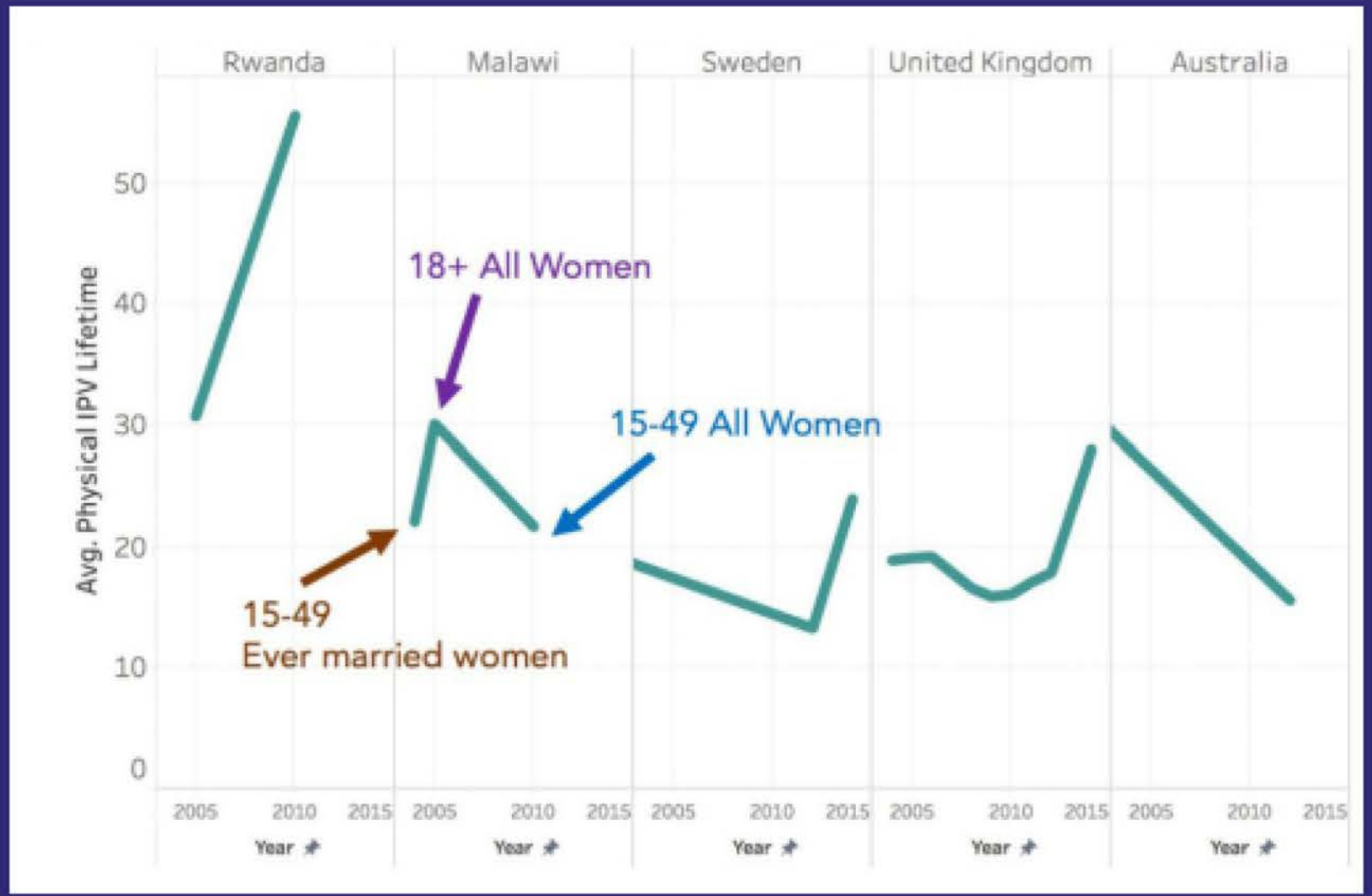
Sample design based on definitions – whose definitions?

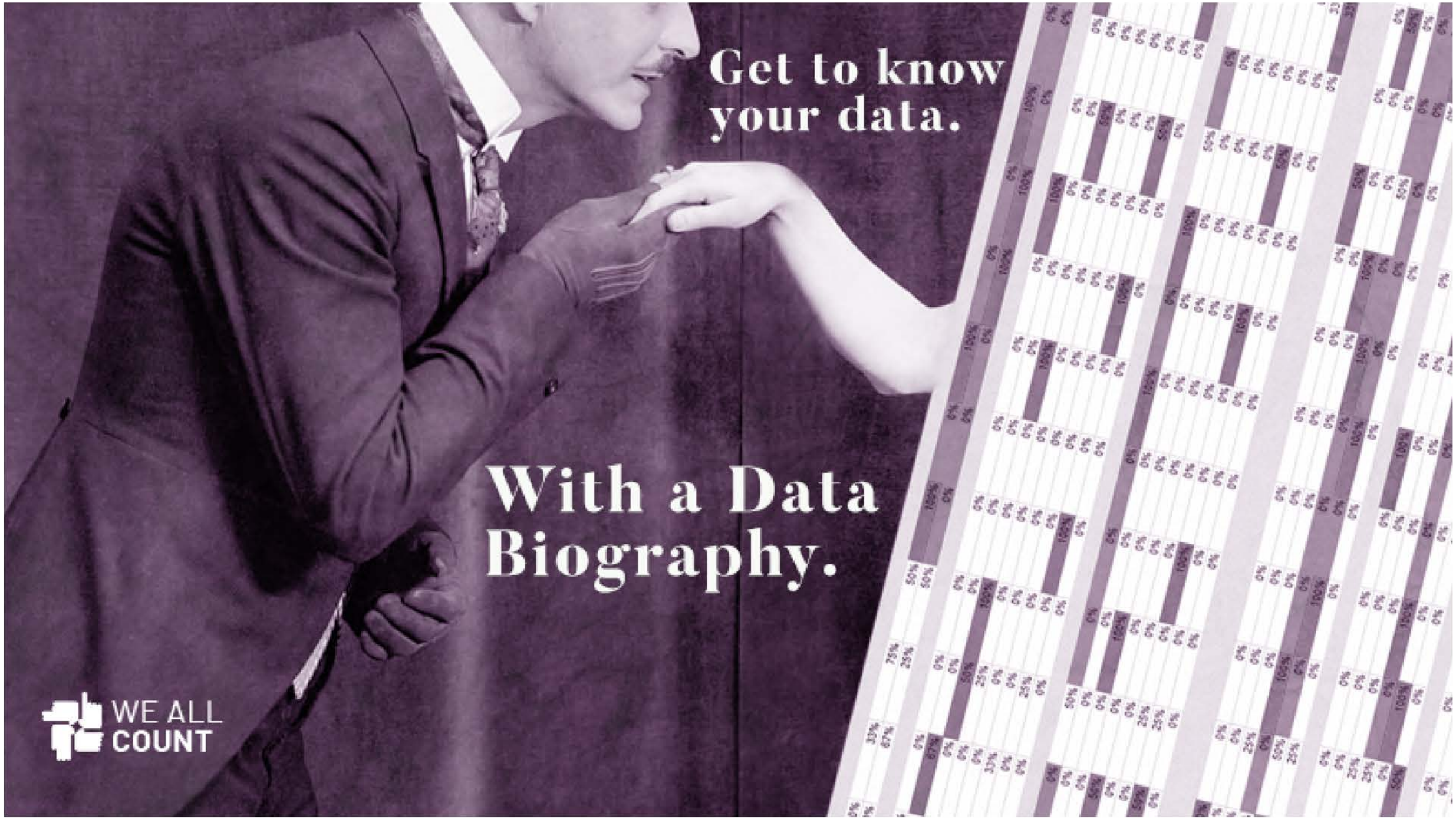


Data Collection & Sourcing




Rwanda
Malawi
Sweden
United Kingdom
Australia
Avg. Physical IPV Lifetime
50
40
30
20
10
0
No Source!!!!
18-64
Ever married women
18-74
2005
2010
2015
Year

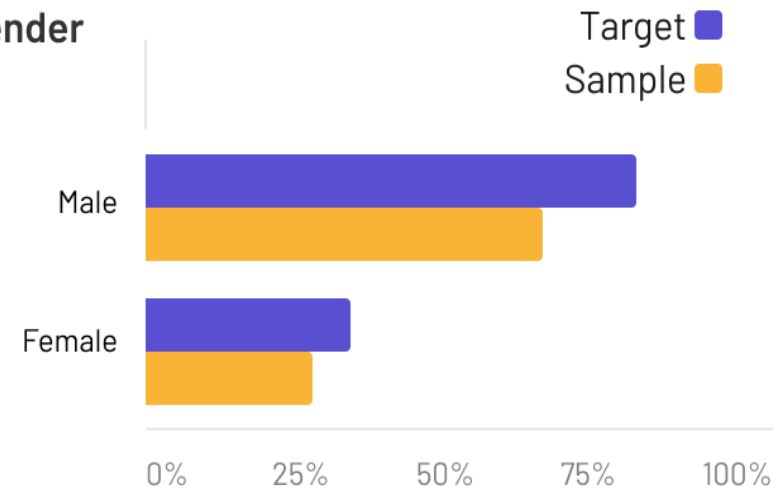

Rwanda
Malawi
Sweden
United Kingdom
Australia
Avg. Physical IPV Lifetime
50
40
30
20
10
0
18+ All Women
15-49 All Women
15-49
Ever married women
2005
2010
2015
Year


Get to know
your data.
With a Data
Biography.
WE ALL
COUNT

```
The short version of the data biography is the basics. It consists of four core questions:
**Who?**
Who collected the data?
Who owns the data?
Who is included within the data?
Who is intentionally or unintentionally excluded from the data?

**How?**
What are the methods used for the data collection design and implementation?

**Where?**
In what locations was the data collected?
Where is the data being stored?

**Why?**
For what specific purpose was each dataset you're using collected?
Do you have informed consent to use the data for any other purpose?

**When?**
When was the data collected?
```

Parameter 1

Gender



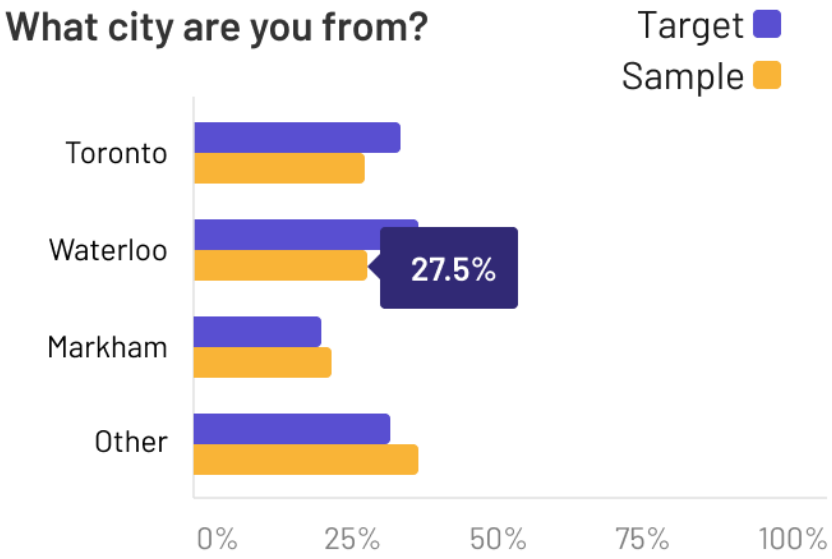
Your results

High Bias | Some Bias | Less Bias

Your results are biased.
Parameter 1 (Female) were surveyed more often than parameter 2 (male).

Parameter 2

What city are you from?



Your results

High Bias — Some Bias — **Less Bias**

Your results are not biased.
All survey parameters were surveyed equally.

```
#Perform test to compare sample and population proportions
keyvar1z <- chisq.test(c(countKeyVar1.1,countKeyVar1.2), p = keyvar1pr)
unclass(keyvar1z)
keyvar1z$p.value

FinalOutputkeyvar1z <- "yellow"
FinalOutputkeyvar1z[keyvar1z$p.value > .05] <- "green"
FinalOutputkeyvar1z[keyvar1z$p.value > .001 &keyvar1z$p.value > .0499] <- "yellow"
FinalOutputkeyvar1z[keyvar1z$p.value < .001] <- "red"
```

Analysis

Methods Matter

A LOT.

What statistical method you use is based on your world view.

Special education intervention to help vulnerable kids read better.

Tested at beginning of year and end of year.

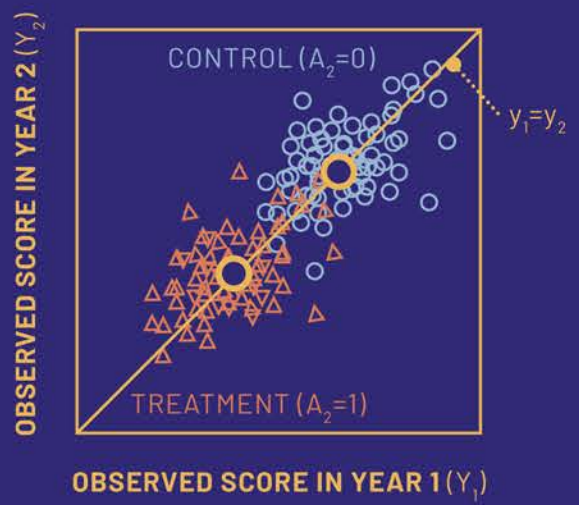
Assign the treatment to kids who are more likely to need help - since funds are limited.

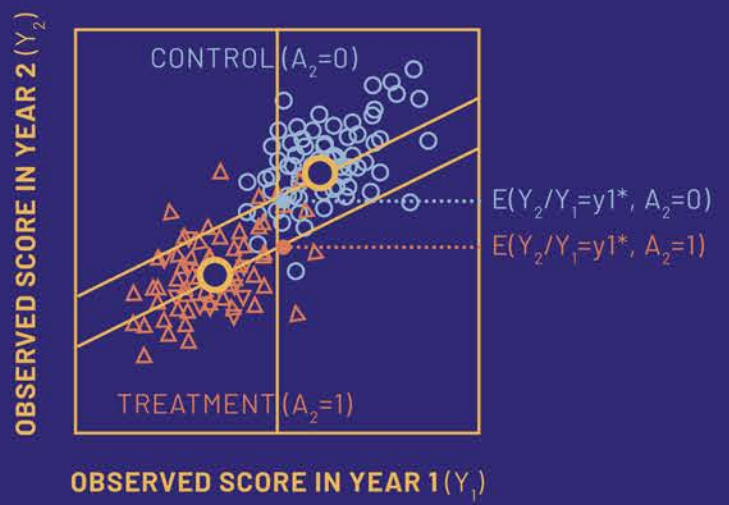
Two analyses done.

One difference in differences, one regression with baseline as covariate.

Analysis #1 concludes that the intervention has no average effect on student reading performance.

Analysis #2 concludes that the intervention has a large negative effect on student reading performance.





Another Example is thinking about statistical models that look at punishment - either in the criminal justice system or in educational discipline settings.

We have reasonably good data on what punishments are handing out - but not on what behaviors actually happened.

So we have isses that are very often accidentally biased.

Ways to think about fair

Equal False Negative Rates: the fraction of positives which are marked negative in each group agree.

Equal False Positive Rates: the fraction of negatives which are marked positive in each group agree.

Equal Positive Predictive Values: the fraction of those marked positive which are actually positive in each group agree.

Statistical Parity (equal positive decision rates): the fraction marked positive in each group should agree.

Ways to think about fair

Balance for the Positive Class: the average score assigned to positive members, E[S | Y=1]E[S | Y=1], should be the same across groups. If the score were 0-1 this would be equal true positive rates(equivalently, equal false negative rates).

Balance for the Negative Class: the average score assigned to negative members, E[S | Y=0]E[S | Y=0], should be the same across groups. If the score were 0-1 this would be equal false positive rates(equivalently, equal true negative rates).

Calibration: the fraction of those marked with a given score who are actually positive, E[Y | S=s]E[Y | S=s], should be the same across groups. If the score were 0-1, this would be equal positive predictive values and equal negative predictive values.

AUC (Area Under Curve) Parity: the area under the receiver operating characteristic (ROC) curveshould be the same across groups. The AUC can be interpreted as the probability that a randomly chosen positive individual (Y=1Y=1) is scored higher than a randomly chosen negative individual (Y=0Y=0).

Accuracy Parity

The *Accuracy Parity* is basically how often the algorithm predicts the correct answer, comparing rates of correctness across groups. To find the Accuracy Parity metric (Friedler et al. 2018) use the following:

```
acc_parity(actuals = df$label_value, predicted = df$score,
           group = df$race, base = "Caucasian")
```

```
##         Caucasian African-American             Asian          Hispanic
##         1.0000000         0.9527275         1.2594662         0.9865416
##  Native American             Other
##         1.1609895         0.9938140
```

Here *actuals* is the vector of actual target variables - the known outcome; *predicted* is the predicted target values - the prediction made by your algorithm; *group* is one selected sensitive or vulnerable group (can be binary or a factor); and *base* is the baseline for comparison.

Demographic Parity

Demographic Parity is a measure of how equal Yes predictions are across groups. To find the Demongraphic Parity metric (Calders and Verwer, 2010) use the code:

```
dem_parity(predicted=df$score, group= df$race, base  = "Caucasian")
```

```
##        Caucasian African-American            Asian         Hispanic
##        1.0000000        0.7597998        1.0980033        1.0497301
##  Native American            Other
##        0.6813366        1.1384542
```

Here *predicted* is the vector of predicted target values; *group* is still one selected sensitive or vulnerable group; and *base* is the baseline for comparison.

Disparate Impact

The concept of *Disparate Impact* is one of the most relied on currently. It measured the degree to which subgroups are disproportionately affected by errors. This measures This function calculates the Disparate Impact Metric (Feldman, et al, 2015)

```
dis_impact(predicted=df$score,group= df$race,base  = "Caucasian")
```

```
##         Caucasian African-American             Asian          Hispanic
##         1.0000000         1.6902240         0.7183841         0.8570987
##  Native American             Other
##         1.9156909         0.6021469
```

Here *predicted* is the vector of predicted target values; *group* is still one selected sensitive or vulnerable group; and *base* is the baseline for comparison.

False Negative Parity

The *False Negative Parity* (or FNR)is the measure of how many times we predict No when the answer is really Yes - compared across groups. For example, does our model say No when it's really Yes more often for Hispanic people than for Caucasian people? These calculations are based on (Chouldechova 2017)

```
fnr_parity(actuals = df$label_value,predicted=df$score,group= df$race,base  = "Caucasian")
```

```
##        Caucasian African-American             Asian          Hispanic
##        1.0000000        0.5864159        0.6984816        1.1651395
##  Native American             Other
##        0.2095445        1.4179701
```

Here *actuals* is the vector of actual target variables - the known outcome; and *predicted* is the vector of predicted target values; *group* is still one selected sensitive or vulnerable group; and *base* is the baseline for comparison.

False Positive Rate Parity

The *False Positive Parity* is the measure of how many times we predict Yes when the answer is really No - compared across groups. For example, does our model say Yes when it's really No more often for Caucasian people than for Black people? False Positive Rate Parity estimates are based on (Chouldechova 2017)

```
fpr_parity(actuals = df$label_value,predicted=df$score,group= df$race,base = "Caucasian")
```

```
##         Caucasian African-American             Asian          Hispanic
##         1.0000000         1.9120926         0.3707487         0.9158867
##  Native American             Other
##         1.5988539         0.6290573
```

Here *actuals* is the vector of actual target variables - the known outcome; and *predicted* is the vector of predicted target values; *group* is still one selected sensitive or vulnerable group; and *base* is the baseline for comparison.

Negative Positive Value Parity

Negate Positive Value Parity (see Aeuquitas bias audit toolkit)

```
npv_parity(actuals = df$label_value,predicted=df$score, group= df$race,base = "Caucasian")
```

```
##        Caucasian African-American             Asian          Hispanic
##        1.0000000         0.9137277         1.2291484         0.9993459
##  Native American             Other
##        1.1706175         0.9804904
```

Here *actuals* is the vector of actual target variables - the known outcome; and *predicted* is the vector of predicted target values; *group* is still one selected sensitive or vulnerable group; and *base* is the baseline for comparison.

Positive Positive Value Parity

This *Positive Positive Value Parity* is also called the Precision. This is the fraction of actual Yes cases correctly predicted to be Yes out of all predicted positive cases, When the answer is Yes, how often do we actual predict Yes - and are our rates the same across sensitive subgroups? (see Aeuquitas bias audit toolkit)

```
ppv_parity(actuals = df$label_value,predicted=df$score, group= df$race,base = "Caucasian")
```

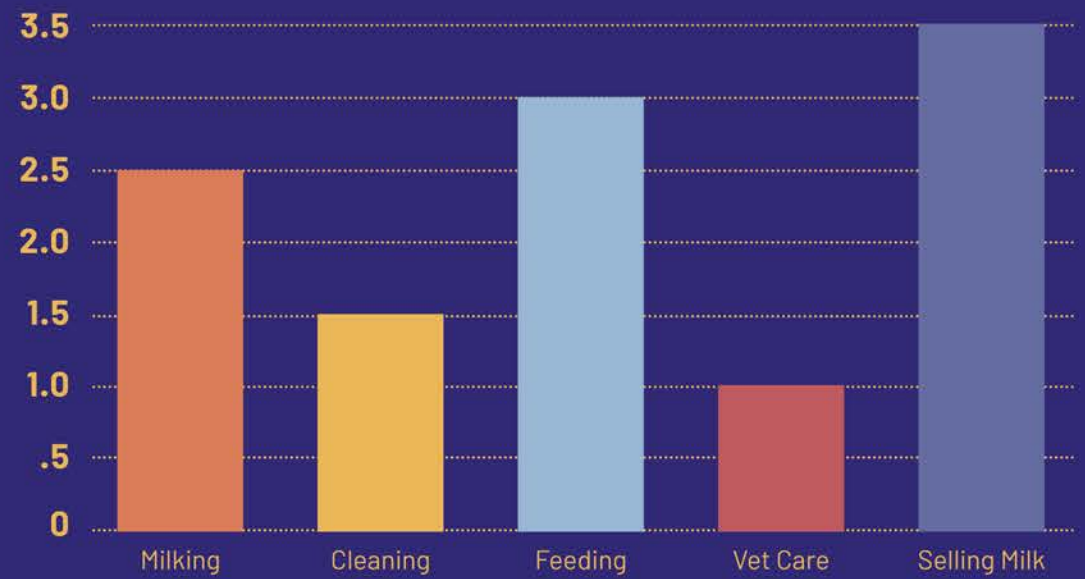
```
##        Caucasian African-American             Asian          Hispanic
##        1.0000000         1.0649039         1.2683168         0.9167483
##  Native American             Other
##        1.2683168         0.9204662
```

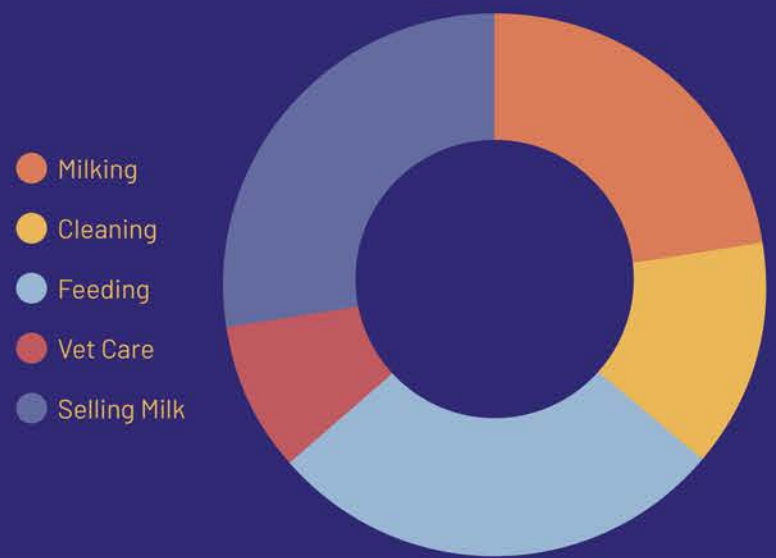
Here *actuals* is the vector of actual target variables - the known outcome; and *predicted* is the vector of predicted target values; *group* is still one selected sensitive or vulnerable group; and *base* is the baseline for comparison.

Interpretation

Communication & Distribution

Data Viz “best practices” are not culturally universaly.





TIME SPENT ON DAIRY ACTIVITIES PER DAY

Sources of bias can be identified in each step of the data life cycle.

Funding

Motivation

Project Design

Data Collection & Sourcing

Analysis

Interpretation

Comunication & Distribution


DATA
DEMYSTIFY. DEMOCRATIZE. DEMONSTRATE.

Thank you.

WeAllCount.com

Heather Krause, PStat

heather@idatassist.com

@datassist