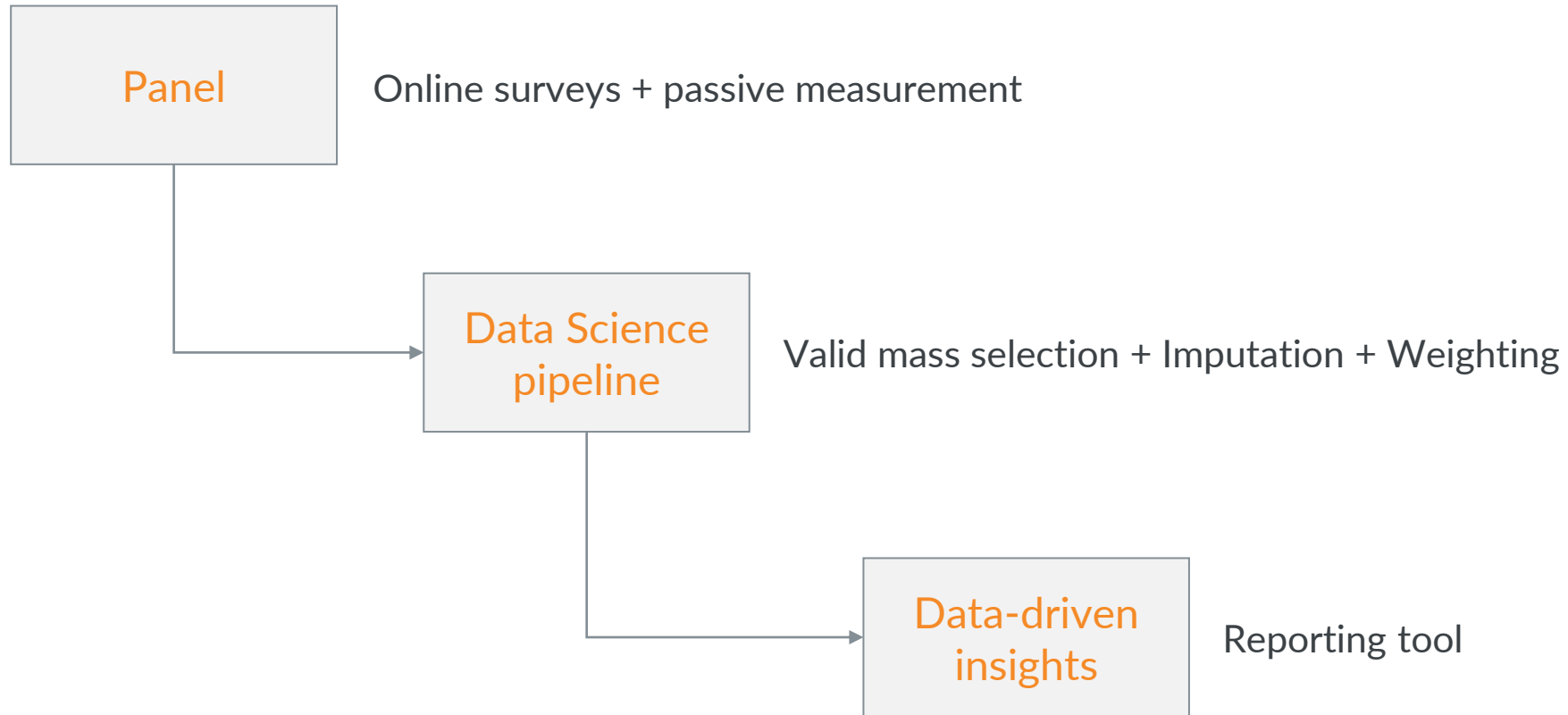




# Goal of the presentation

From panel data acquisition to insights on digital consumption:  
Data needs/difficulties to deliver insights



---

# Panel

## Panelists

willing to participate in an online survey  
give away digital consumption

## Points

as incentives (online shop)

## Recruitment

social media  
mailing lists  
MGM

# Panel

Overview on data acquired



## What do we know about panelists?

### Sociodemographics

Age, gender, region, educational level, working sector, number of kids...

### Reported device usage

Device types panelists report to use to access Internet

### Digital passive measurement

Panel meter technology – digital consumption  
Desktop, Smartphone and Tablet

# Panel

Use of survey data - examples



What can  
we learn  
from digital  
surveys?

Which online site do you would visit if you were to buy a new pair of sneakers?

**Target: female – aged 30-45**

---

Which party will you vote in the next country elections?

**Target: overall population**

---

How many cans of soft drinks do you consume per week?

**Target: population aged 15-35**

---



# Panel

Use of meter data - examples



What can  
we learn  
from digital  
traffic?

Reach share of online sites which are visited to buy a pair of sneakers (\*)

**Target: female – aged 30-45**

---

Most read online newspaper

**Target: aged 65+**

---

List of top 20 gaming apps

**Target: population aged 15-35**

---

---

## Data Science Pipeline

### Valid mass

selection of usable panelists

### Imputation

completeness of panel data

### Weighting

tackling bias

### Imputation

Machine Learning algorithm to **replace missing values** with substitute ones

Why do we impute?

**Make the most** of panel data

Address **underestimation** of KPIs

Panelists may not be willing to give away some information, and are not forced to install the panel meter on all the devices they use



### Imputation

Issue most commonly addressed:

imputation over missing variables (e.g. household income)

Examples of underestimation or bias when using digital consumption:

- **Dropping panelists with gaps in measurement could lead to a systematic error**
  - If these panelists have a specific profile, we will be biasing the panel in that direction
  - Lower coverage
- **Keeping only the tracked data can also lead to systematic errors**
  - If people tend not to meter their tablets, usage on tablet would be underestimated
  - If people meter only their PC but have most usage on smartphone, their usage would be underestimated

### Imputation

Imputation as a **2-step process**:

sociodemographic and questionnaire data  
devices which are reported to be used

Female 35	Device	Claimed usage	Measured usage	Donor / Recipient
	PC	yes	yes	donor
	Smartphone	yes	yes	donor
	Tablet	no	no	no

Complete usage of smartphone is donated

Female 35	Device	Claimed usage	Measured usage	Donor / Recipient
	PC	yes	yes	donor
	Smartphone	yes	no	recipient
	Tablet	no	no	no

Reported device types used to guide imputation step

### Imputation

Imputation as a **2-step process**:

- sociodemographic and questionnaire data
- devices which are reported to be used

In the first step, we make sure there are no missings over:

- reported usage of devices
- drivers to choose donor-recipient pairs

# Data Science Pipeline

Completion of the panel



## Imputation

*There will  
always be  
measurement  
gaps*

Measurable device

There is traffic we will **never be able to track**

public devices

not owned

professional use

**Reported device usage** must only cover **when a device is measurable**

No imputation base for unmeasurable devices

### Weighting

Each **panelist has a weight** so that  
**weighting constraints** are achieved

In one of our projects, weighting dimensions are

share of *gender*

share of *age groups*

share of *region*

share of *device type usage*

# Data Science Pipeline

## Tackling bias



In one of our projects, weighting dimensions are

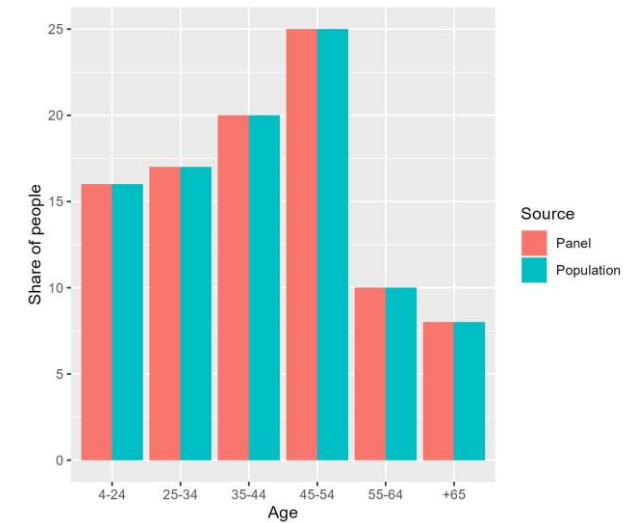
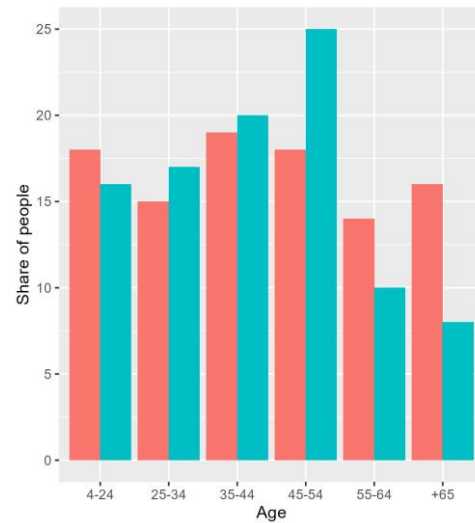
share of *gender*

share of *age groups*

share of *region*

share of *device type usage*

### Weighting



We need these variables to be informed...

# Data Science Pipeline

## Tackling bias

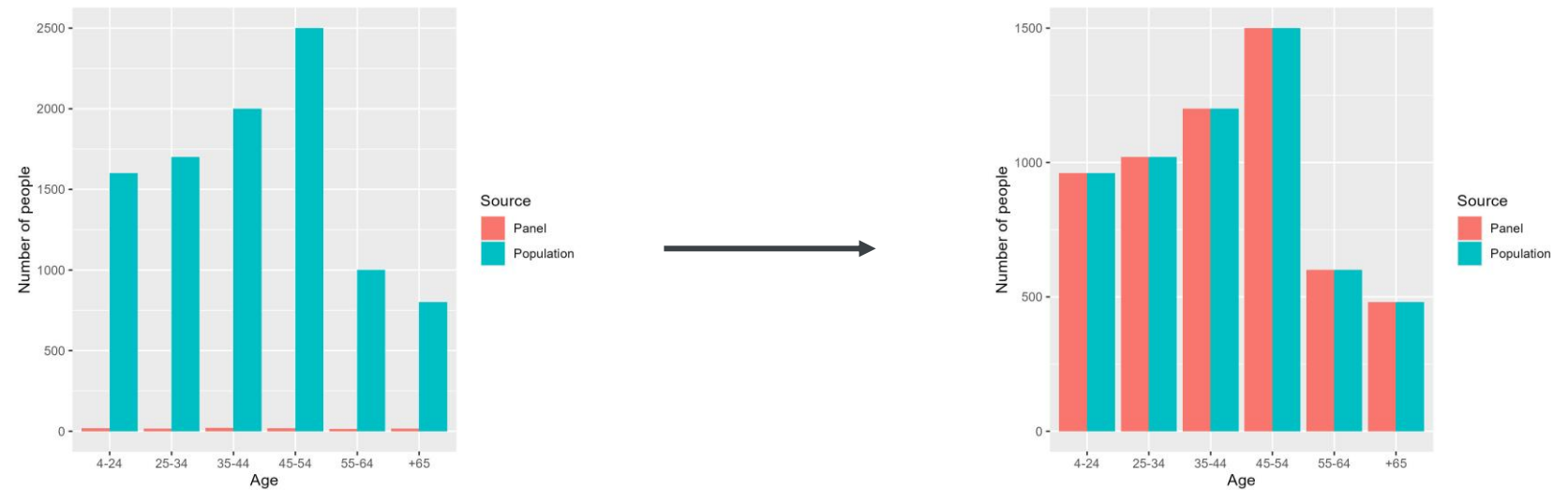


If we seek for insights assuming our panel is representative of a country...

The size of the panel is way smaller than the overall population

### Weighting

Weights can be seen as a **bias correction factor** multiplied by an **extrapolation factor**





# Data Science Pipeline

Selecting the usable panelists



## Valid mass

Selection of panelists who have minimum data to be able to process them...

### Imputation

Informed age, gender, region, household size to impute remaining hooks

At least one validly tracked device

### Weighting

Accepted tolerance for reported device usage (imputed)

A **validly tracked device** must have  
Minimum traffic per month

Difficulty in finding correlation reported frequency of use and actual frequency of use

---

## Reporting tool

Several KPIs to measure our interests

reach  
pageviews  
duration  
sessions  
purchases...

# Reporting tool

Data-driven insights



From our imputed and weighted panel we can obtain...

## Reach

weighted number of panelists

## Pageviews

weighted number of observations

## Purchases

weighted number of purchase URLs

among others...

In combination with targeting over sociodemographic groups

# Takeaways

**Passive measurement** for better understanding of digital consumption VS top-of-mind from surveys

Use of digital surveys may serve not just for answering specific questions on market-related matters, but also to **adjust a panel**

Measurable device and a validly tracked device concepts to understand **tracking coverage**

Digital metering data can be used to **learn about digital consumption** for a specific target via a sufficiently solid method (tackling over/under-estimation errors and potential biases)

Need for a **definition of minimum data requirements** to achieve a solid method

**Thank you!**  
**Questions?**