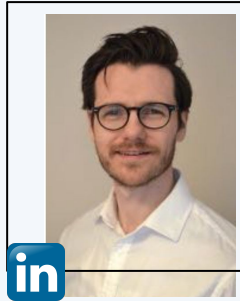


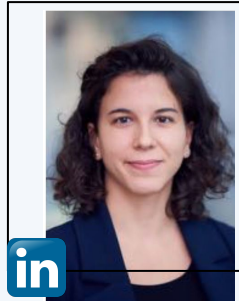


# Meet the team

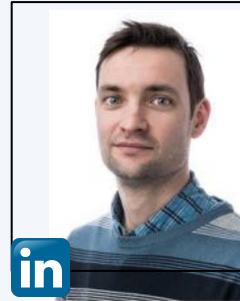
---



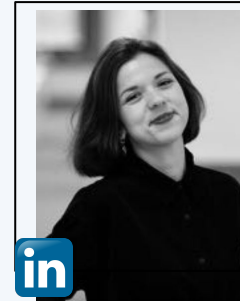
[JAMES DOHERTY](#)



[LORENA GARCIA](#)



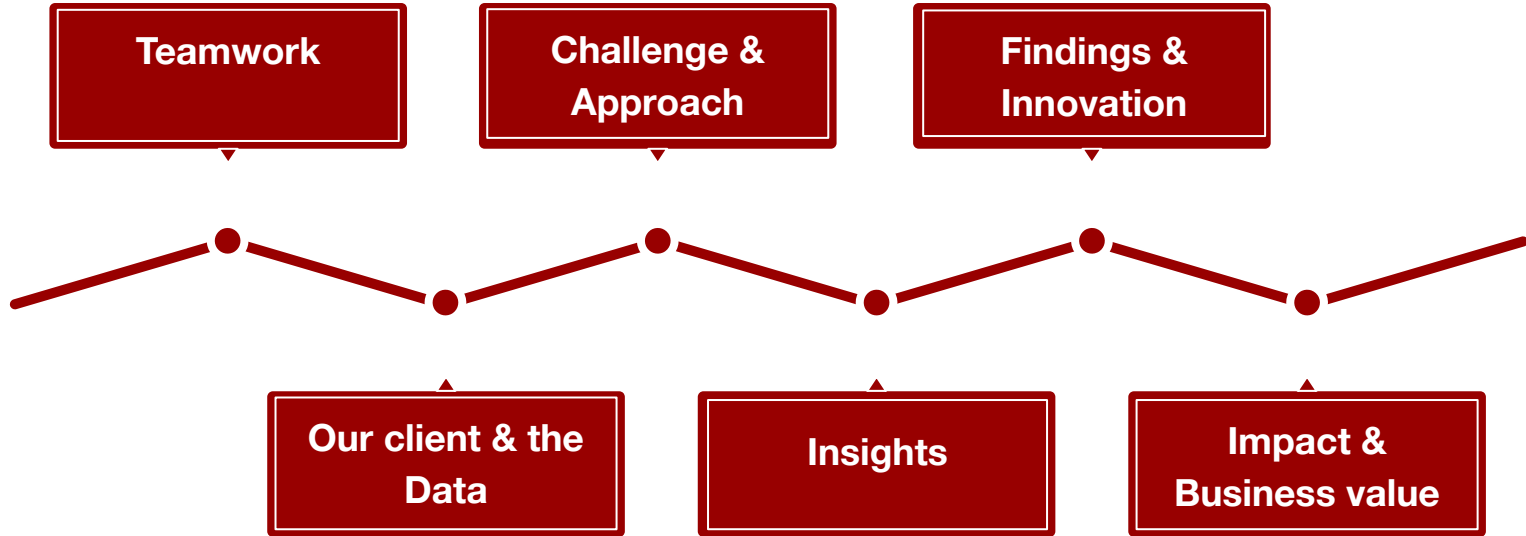
[CHARLIE JEYNES](#)



[MISHKA NEMES](#)

# Roadmap of the next 30 minutes...

---



# Teamwork

---





*Our client*

# Food Standards Agency

---

- Non-ministerial government body
- Responsible for food safety and hygiene in England, Wales and Northern Ireland (NI), and food labelling policy in Wales and NI
- Mission: trust the food you eat and it is what it says it is



A large white circle is centered on a solid orange background. Inside the circle, the text "The challenge" is written in an orange, italicized serif font.

*The challenge*

# The challenge: Wales

---

- **Sampling**

- Do FSA surveys undersample key demographic or socio-economic groups in Wales?

- **Food-related behaviours**

- Do those groups have associated food safety risks?
- What is their understanding of food labelling?
- Should they be directly targeted by public engagement work?

- **Predicting food risk**

- What correlations are present within FSA survey data?
- Can we build a model that predicts food risk?







*Data*

# Datasets

---



- Biennial surveys from **2010-2018**
- **1700 respondents** from Wales
- **14 demographic** questions (e.g. Age)
- **500 questions** on food-related behaviour
- **Categorical & ordinal** data
- Conducted by face-to-face interview



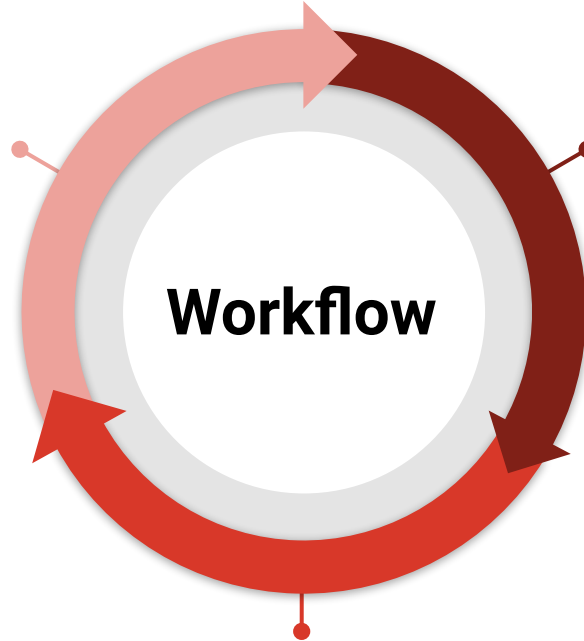
- Most recent census from **2011**
- Data from **53,000 respondents** in Wales
- **120 demographic** questions
- **Categorical & ordinal** data
- **Ground-truth** demographic proportions

A large white circle is centered on a solid red rectangular background. The circle is perfectly circular and occupies a significant portion of the frame.

*Our approach*

---

**Risky food  
behaviour by sub  
demographics**



**Identifying  
undersampled  
groups**

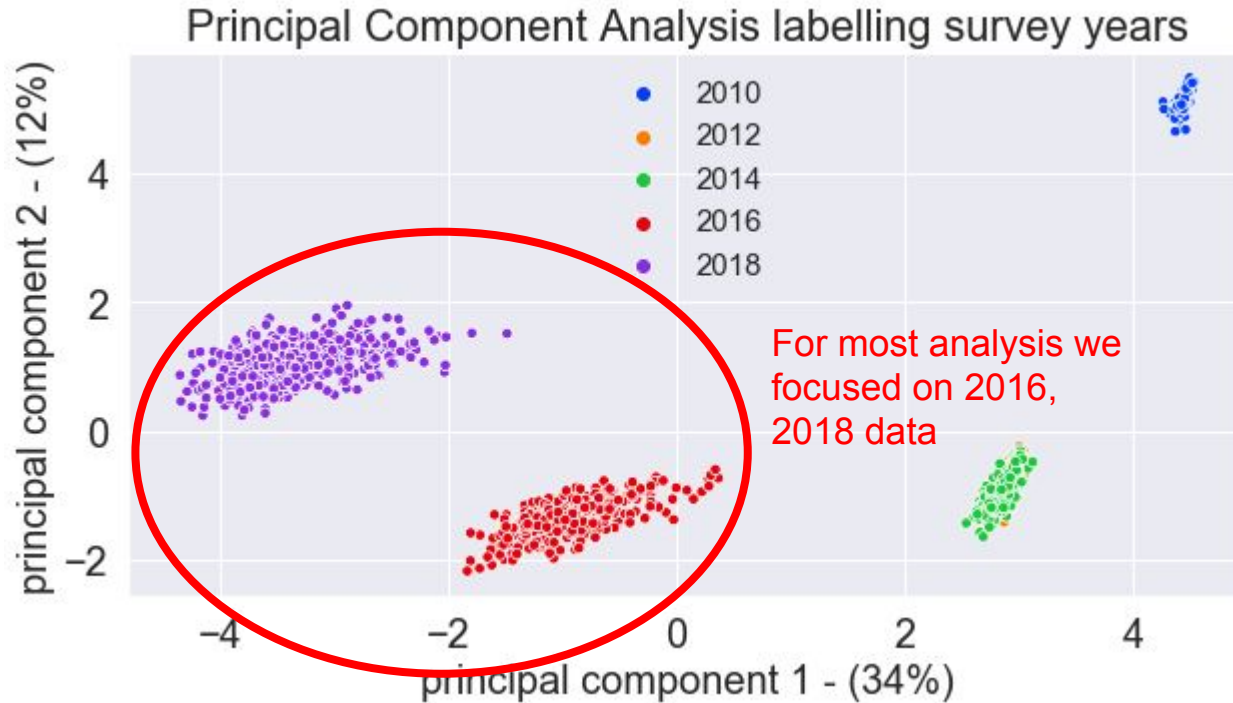
**Quantifying  
uncertainty**

A large white circle is centered on a solid blue rectangular background. Inside the circle, the word "Insights" is written in a blue, italicized serif font.

*Insights*

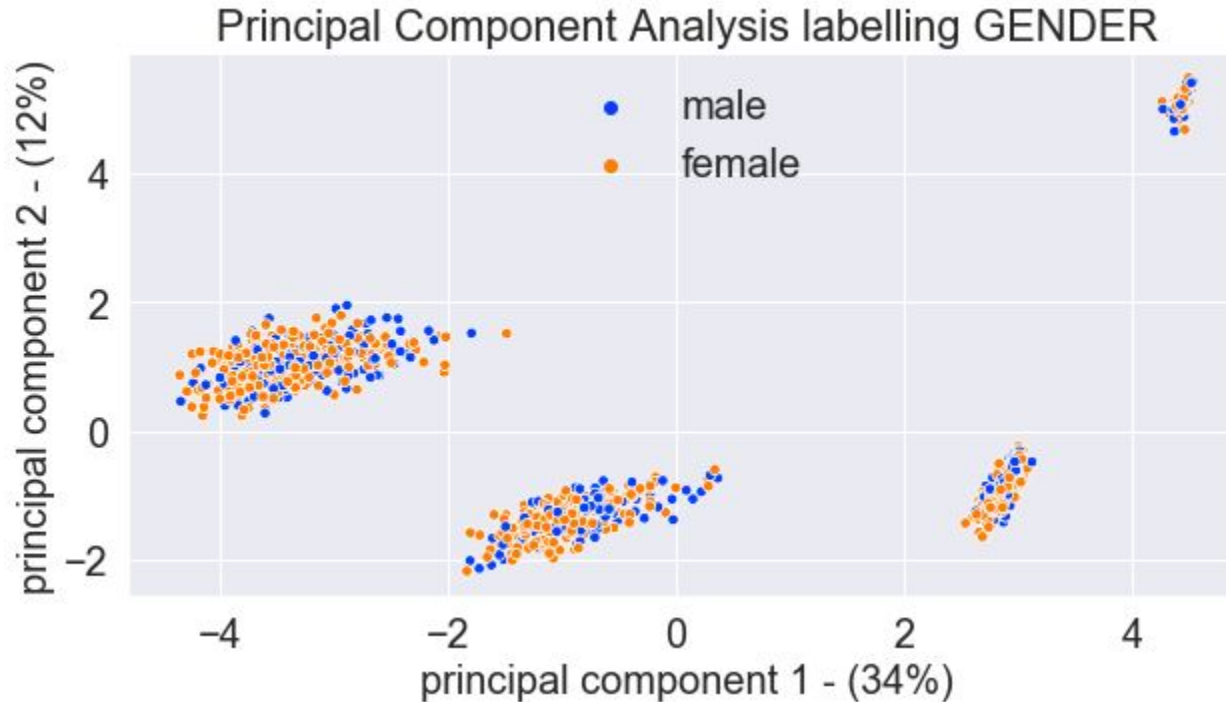
# Differences between **years** in survey: Wales

---

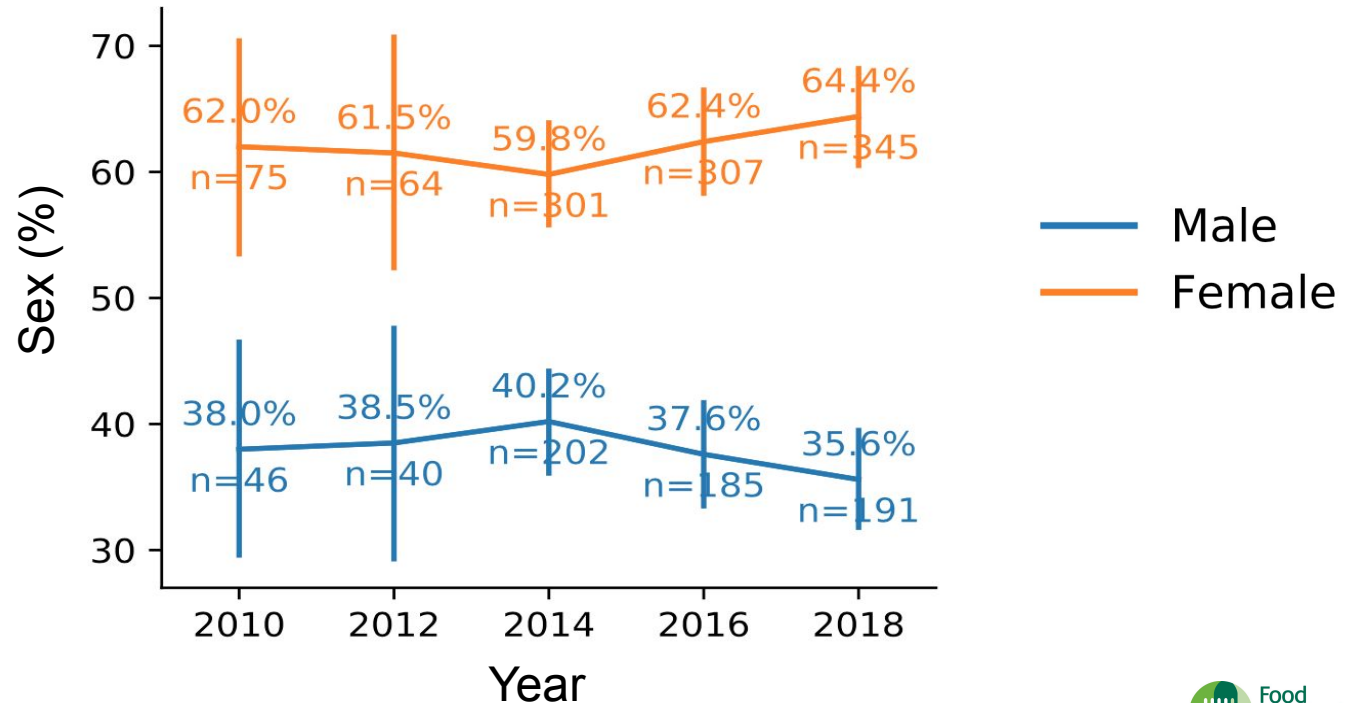


# Differences between gender in whole survey?

---



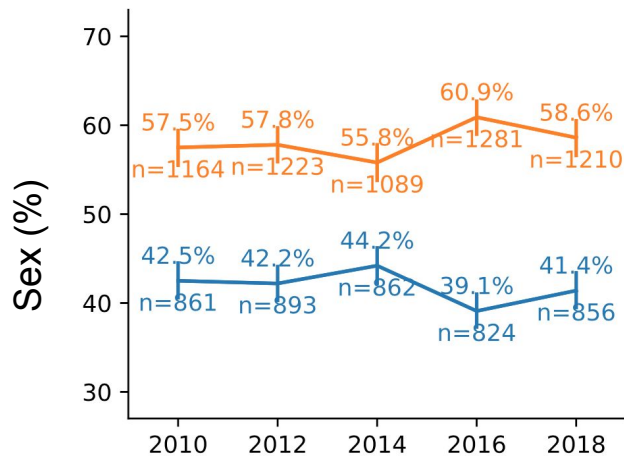
# Gender of F&Y respondents in Wales



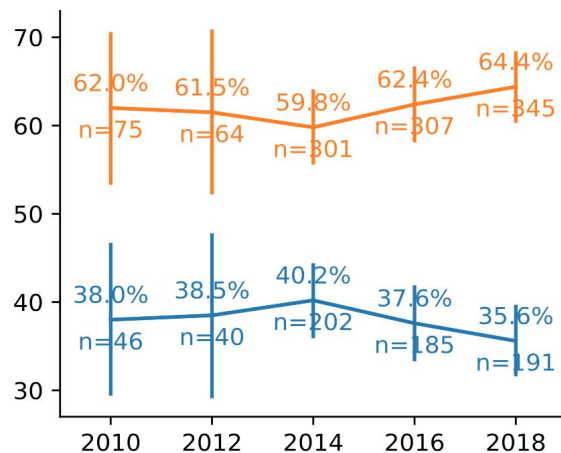


# ... compared with other countries

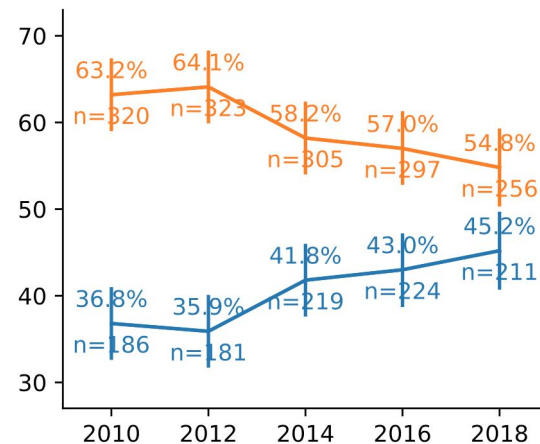
England



Wales



Northern Ireland

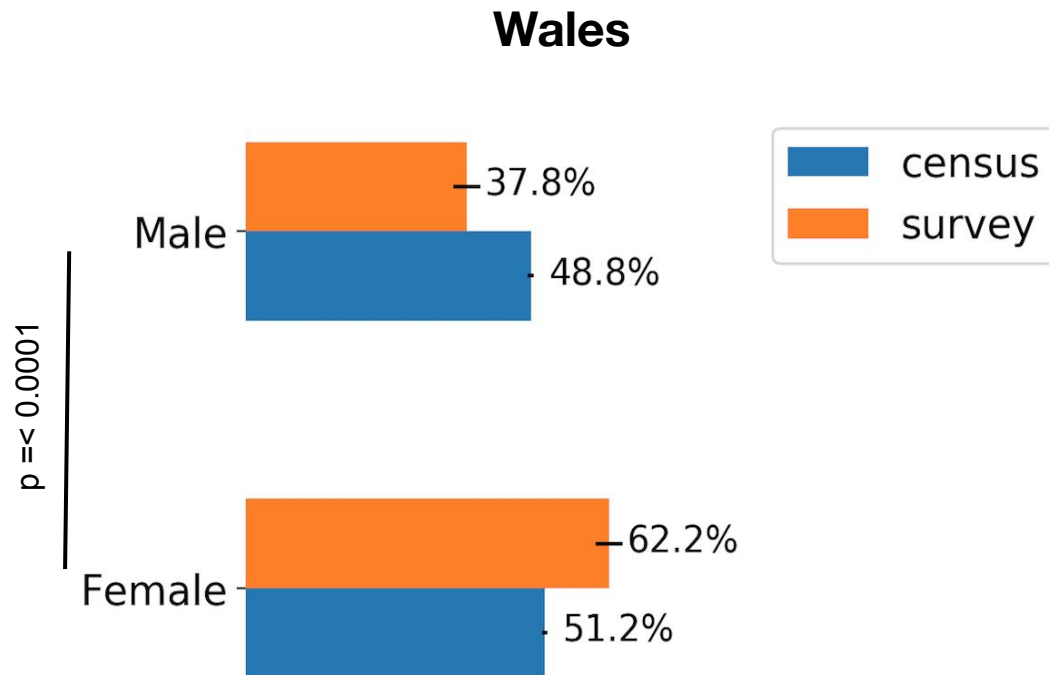


Year

Male  
Female

# Census vs survey respondent gender: Wales

---

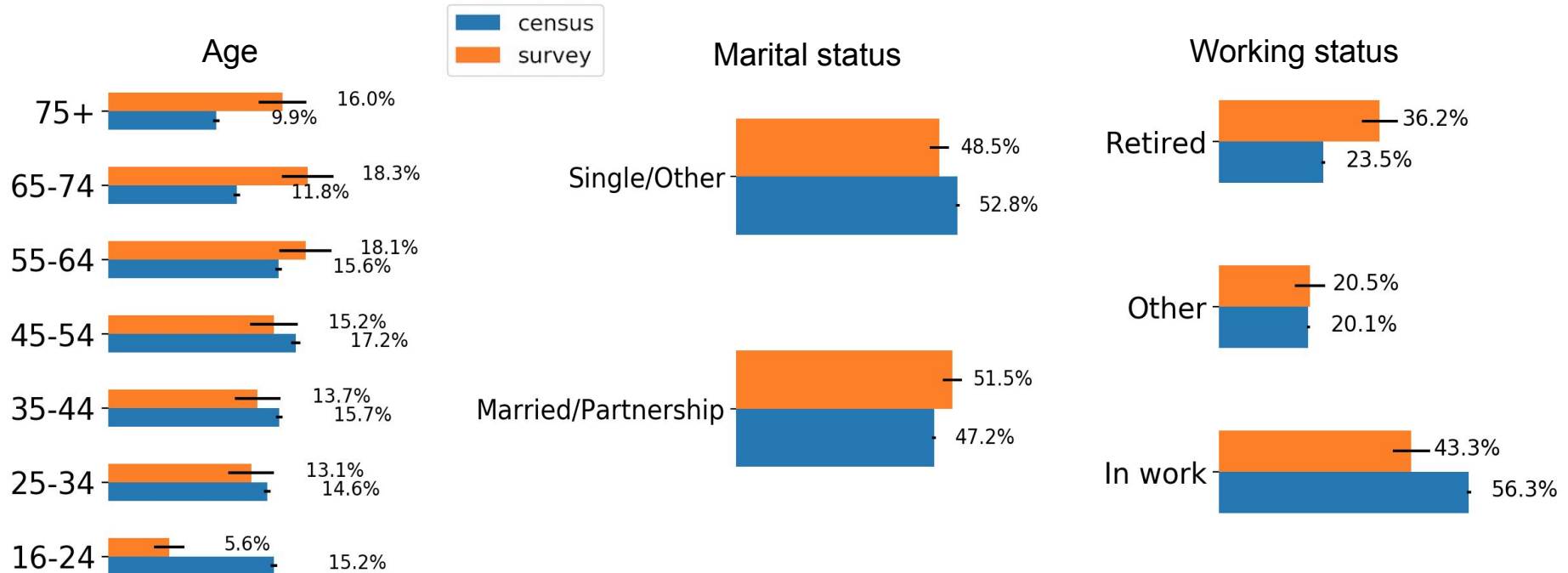


# F&Y demographic information

---

- Gender
- Age
- Household size
- Age of children
- Marital status
- Household income
- Working status
- Region
- Religion
- Deprivation
- Urban-rural
- Health status

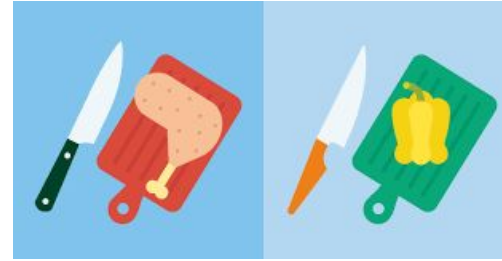
# Census vs survey demographics: Wales



# Which behaviours relate to **food safety**?

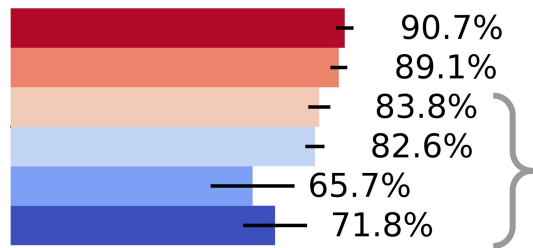
---

Features related to food safety, based on previous **common knowledge**...



# Food safety by demographics: UK wide

Respondents who reported **NO** food poisoning, in the last year

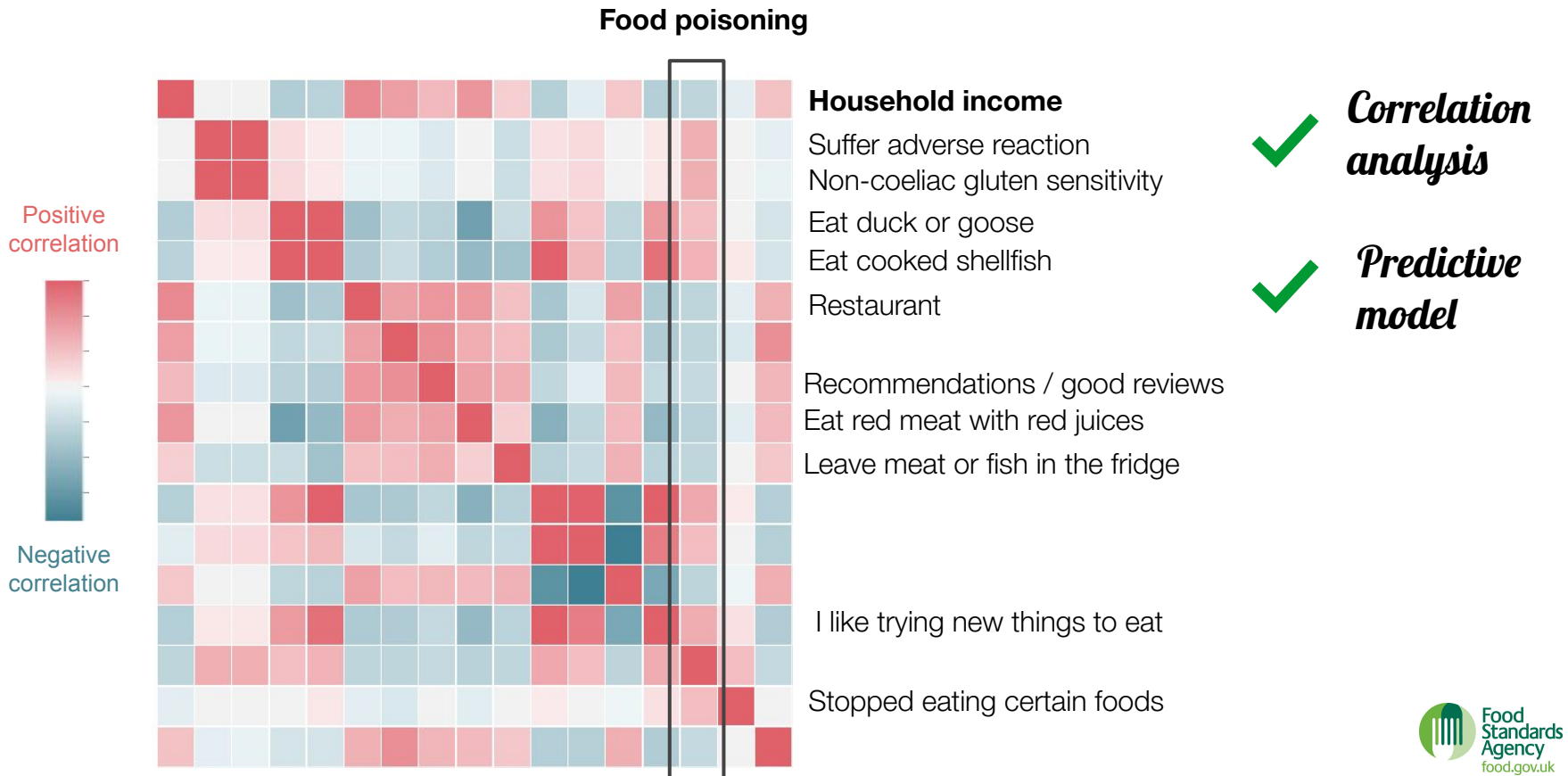


- 55+ Male
- 55+ Female
- 25-54 Male
- 25-54 Female
- 16-24 Male
- 16-24 Female

*Analysis of other features says that...*



# Data-driven features related to food safety: UK



# Interactive web dashboard

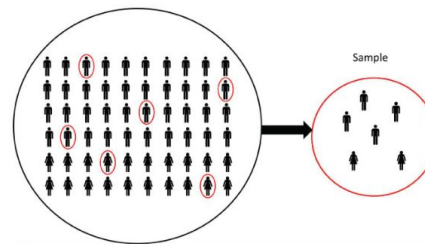
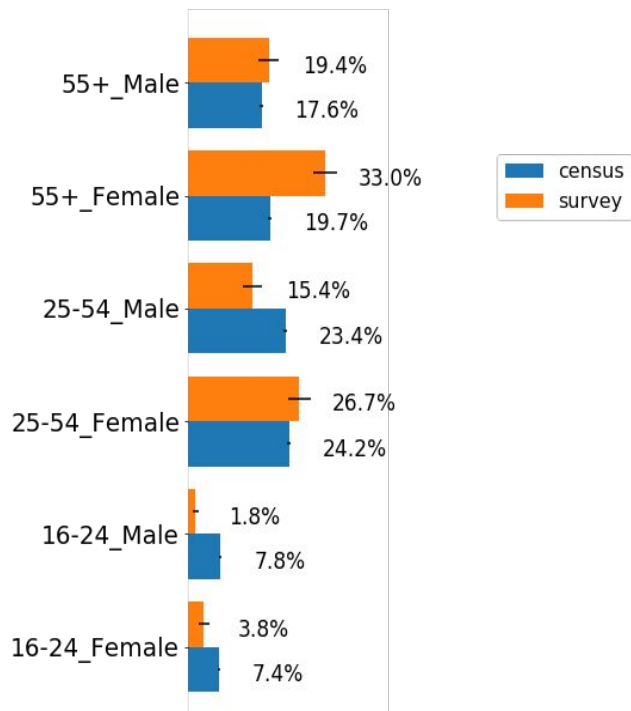
Food risks associated with demographic sub-populations in Wales can be viewed on demand with this [web dashboard](#)



A large white circle is centered on a solid yellow background. Inside the circle, the word "Impact" is written in a dark gray, italicized serif font.

*Impact*

# Missampling in the **real world**: Wales



The real numbers:

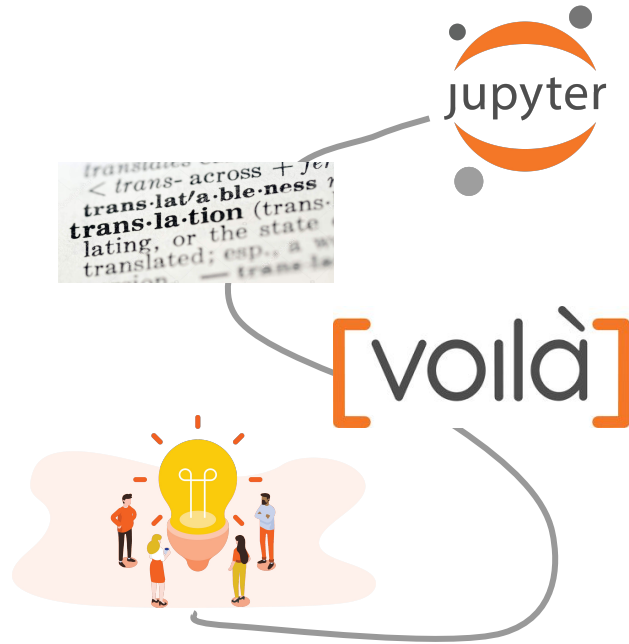
~ **341k** young people (16-24)  
undersampled

~ **557k** females oversampled

(Population of Wales in 2018: 3,187,203)

# Innovation & deliverables

---



- Provided code in a visual and user friendly **Jupyter notebook**
- Created a **dictionary tool** to label data in an automated manner
- Built an **interactive dashboard** for analysing survey data on demand
- Derived insights into survey **demographics** and associated **food behaviours and risks**

*A budding **predictive model** for food poisoning risk*

## Findings



## Recommendations



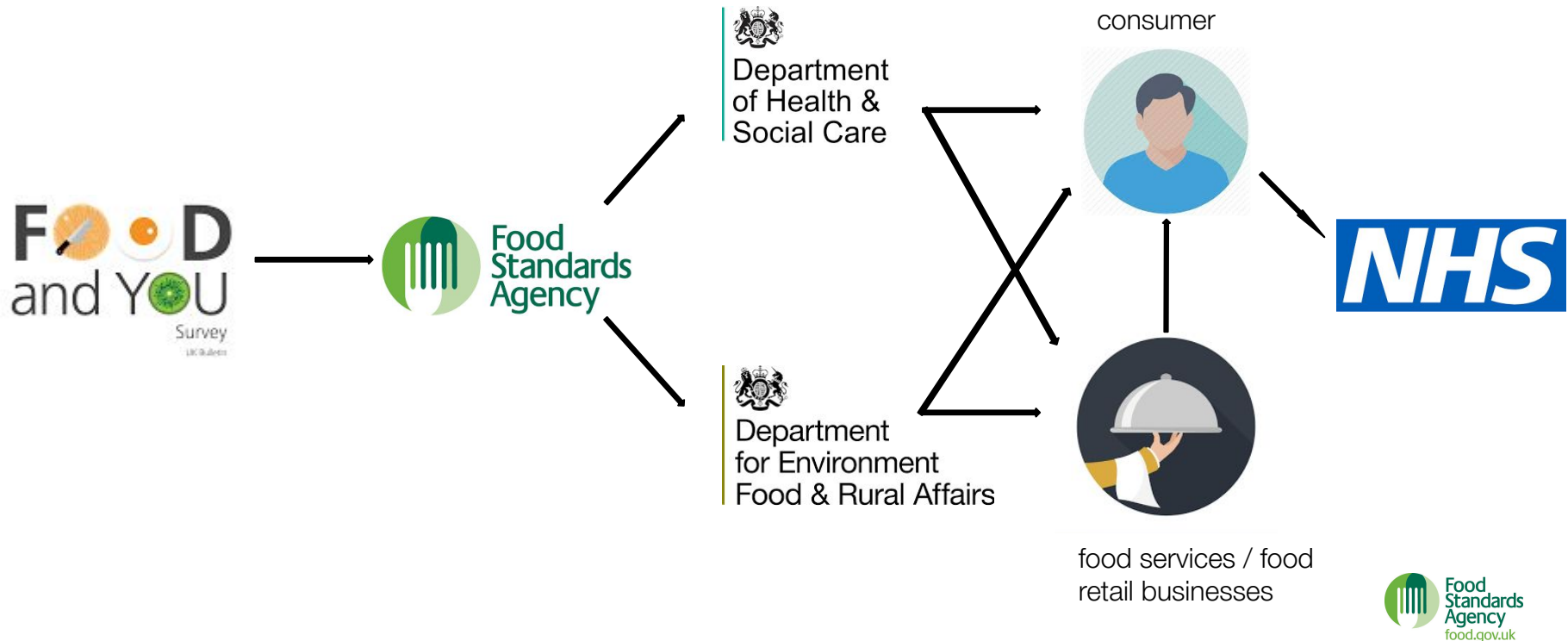
## Business value



<b>Visible trends</b> in the data	Interrogate other questions of interest	Accessible tool that can be queried for any question and demographic
Young people are <b>undersampled</b>	Actively target these groups	Improve food safety for 1/3 million young lives in Wales
Females 55+ are <b>oversampled</b>	Check again with 2021 census	Remove data biases
Missing <b>key demographics</b> & lack of <b>granularity</b>	Redesign the survey demographic options	Better understanding of food behaviour in subpopulations
<b>Bespoke measures</b> for demographics	Use standard measures e.g. as recommended by ONS	Allow direct & reliable comparison with other data sources

# Policy and Public Health Impact

---



Stuart Rylance  
Tom Harley

Andras Szabo  
Maryam Qurashi  
Jon Leslie

***Thank you!***

Wider Pivigo and FSA teams  
Fellow March 2020 S2DS-ers  
Stackoverflow Community



*“The population of sheep in  
Wales is four times greater  
than the Welsh population of  
humans”*

ukpopulation.org