Barcelona GSE Graduate School of Economics

MASTER PROJECT

Measuring real broadband speeds using crowdsourcing data from the Internet Foundation

Author: Ivan Vallejo Vall

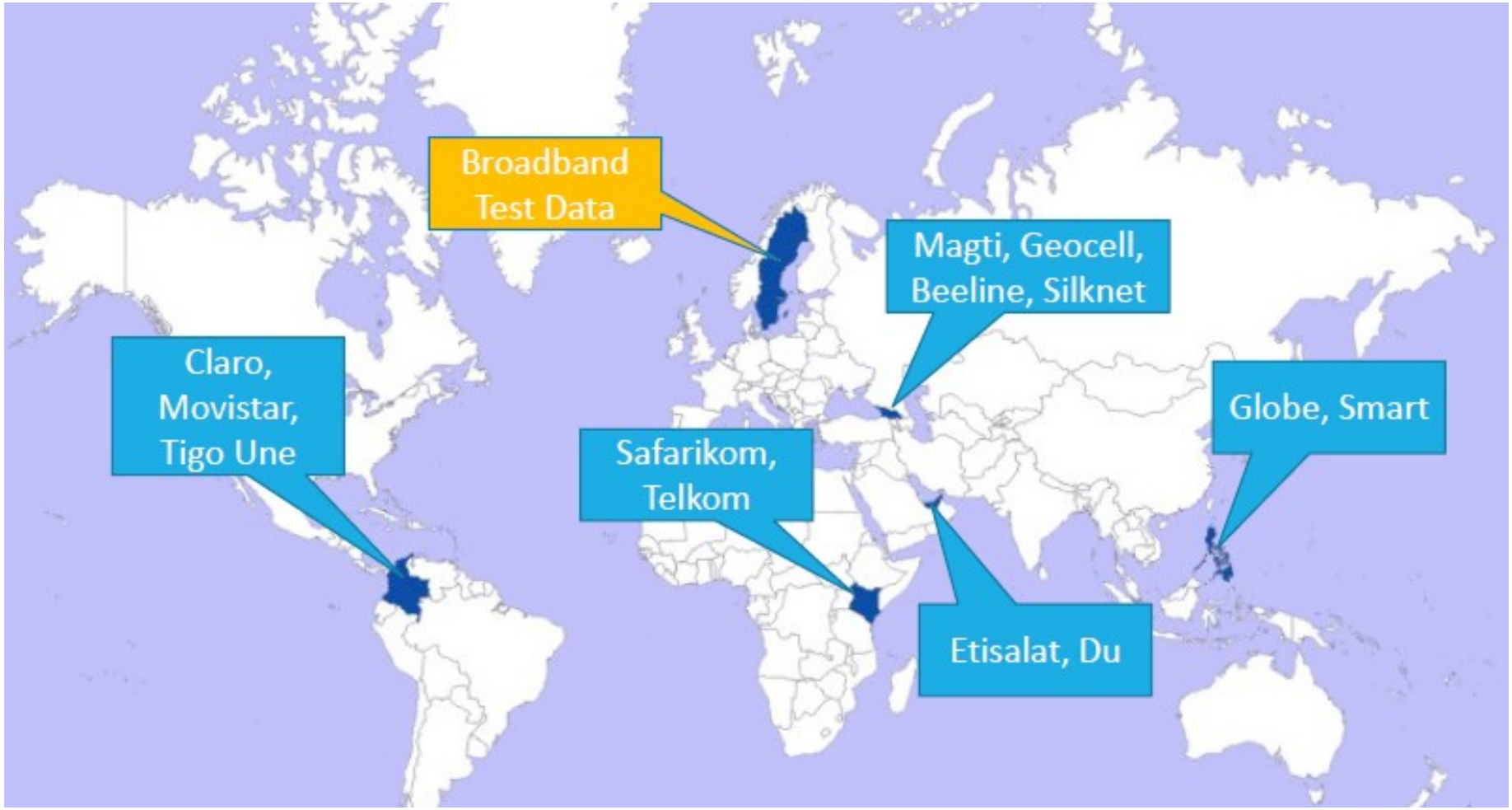
Tutor: Iñigo Herguera García

Data Science Program 2016/2017

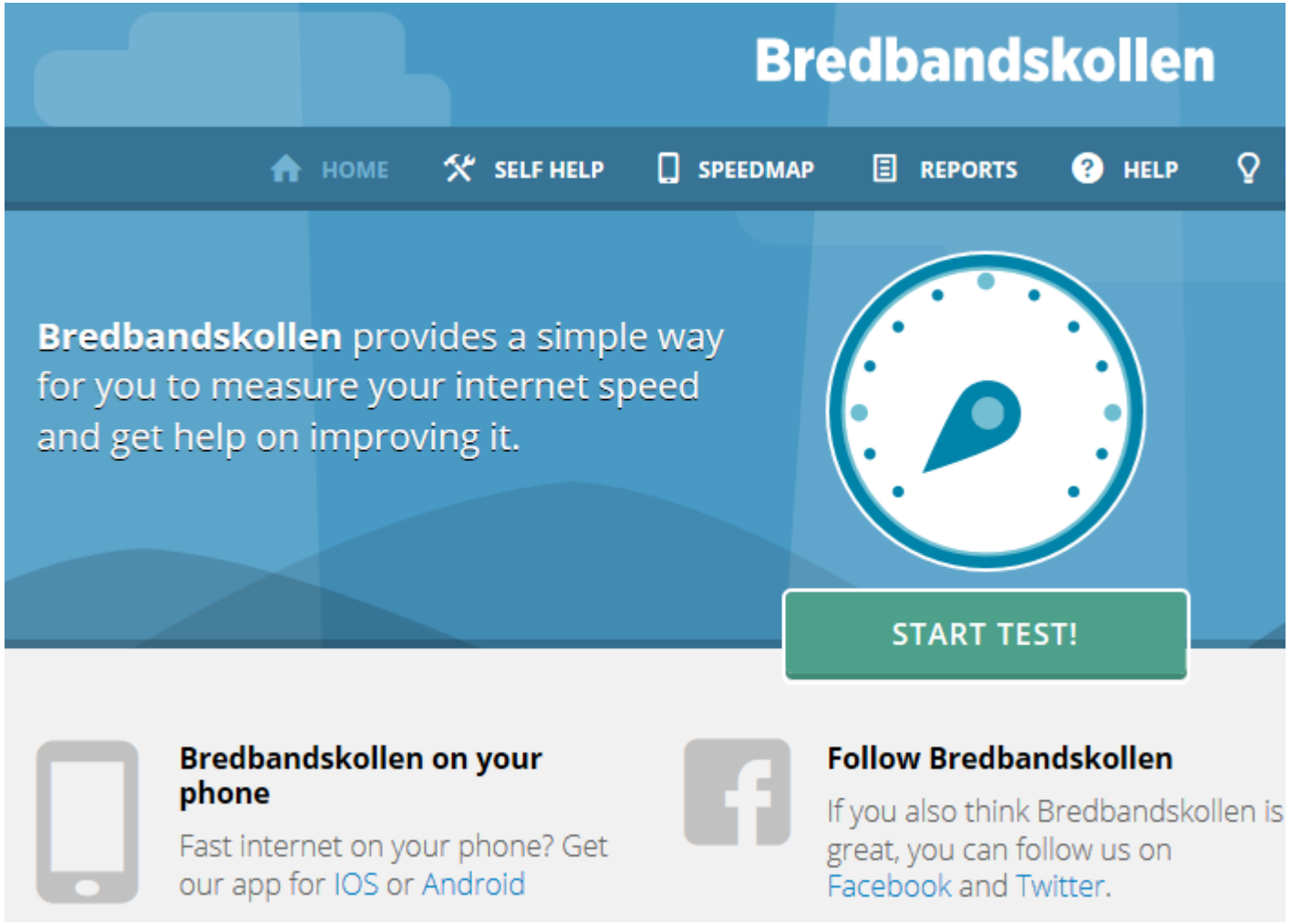
30 June 2017

Index

Big Data for Measuring the Information Society



Bredbanskollen



Relevance of Internet speed

- **Consumers**
  - 2nd most important factor
  - advertised ≠ real speed + congestion [1]
- **Regulators**
  - Quality of service (QoS) [2]
- **Policy makers**
  - Universal Service Funds, broadband plans [3]

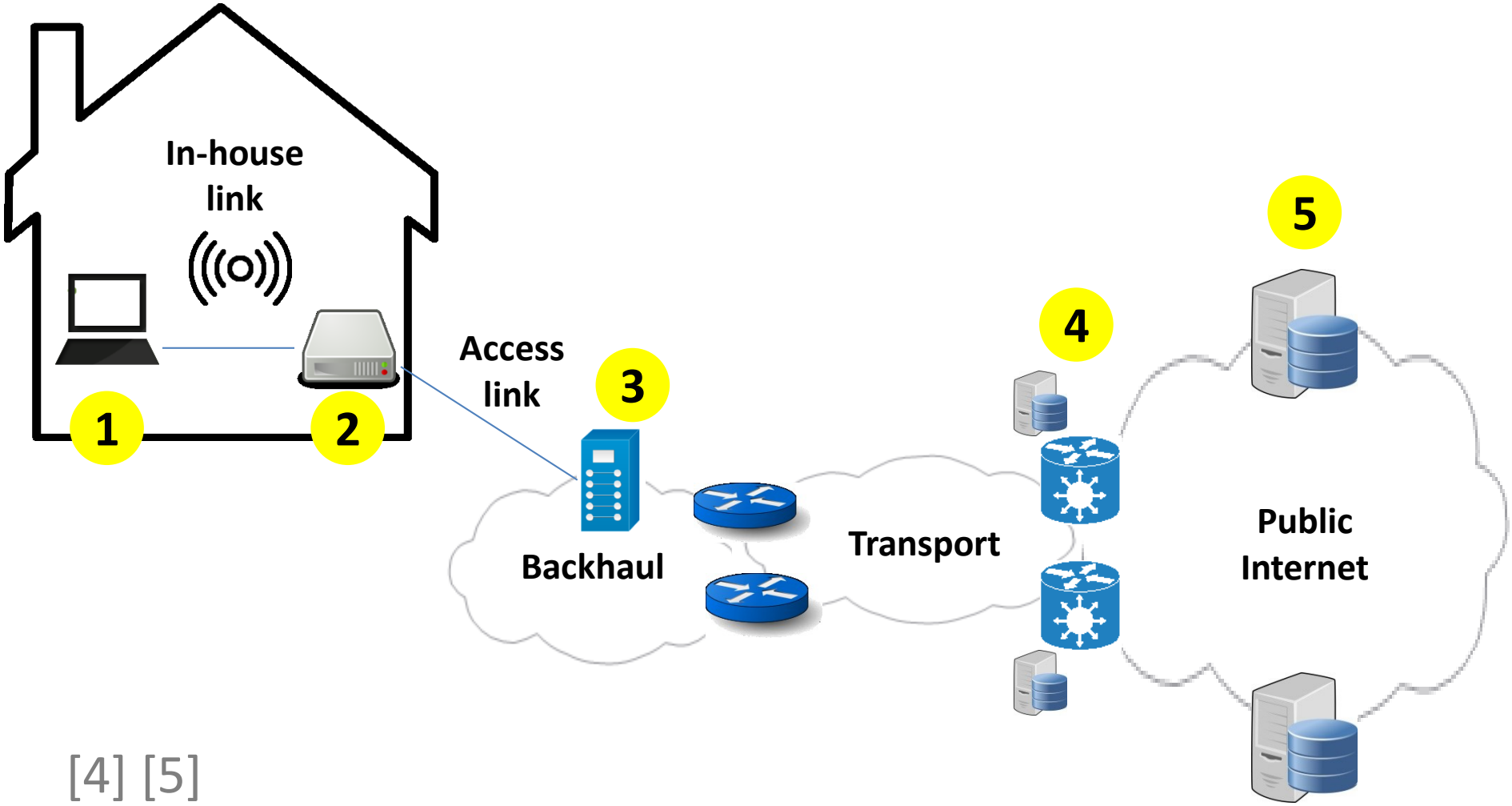
Research question

Can crowdsourcing Internet data be used to measure real broadband speeds?

- Advertised speeds ✕
- Hardware-based measurements  $ $
- Software-based crowdsourcing?

Index

From where to where?



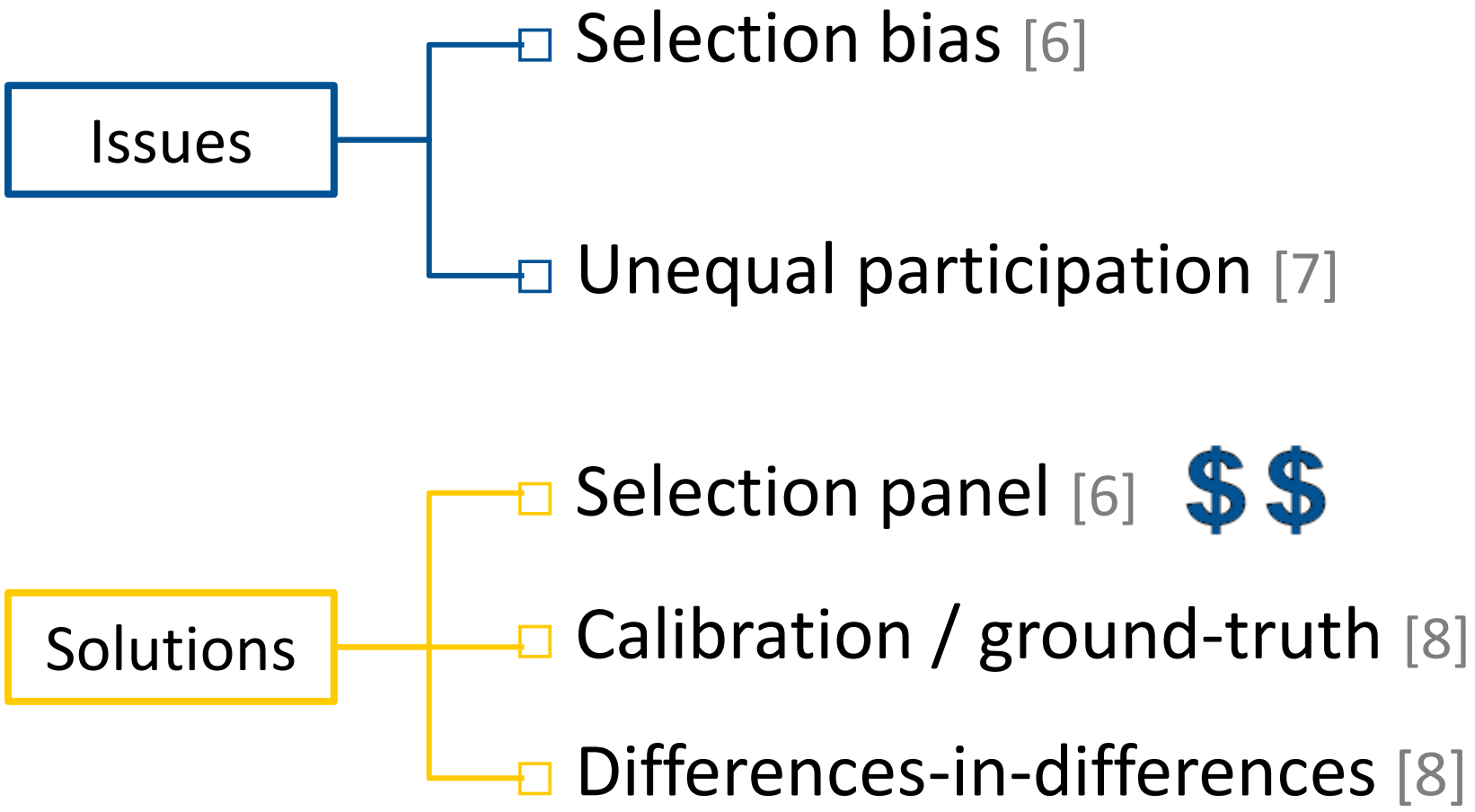
[4] [5]

Speed test comparison

	Test path	Active Passive	Voluntary Automatic	Statistical aggregation
Akamai	1-4 or 1-5	Passive	Real traffic	Unknown
Ookla	1-4 or 1-5	Active	Voluntary	Avg. after ramp-up. Excl. top 10% & bottom 30%
Breadband skollen	1-4 or 1-5	Active	Voluntary	Avg. first 2s or avg. 10s
SamKnows	2-5	Active	Automatic	Avg. after ramp-up. Excl. top & bottom percentiles

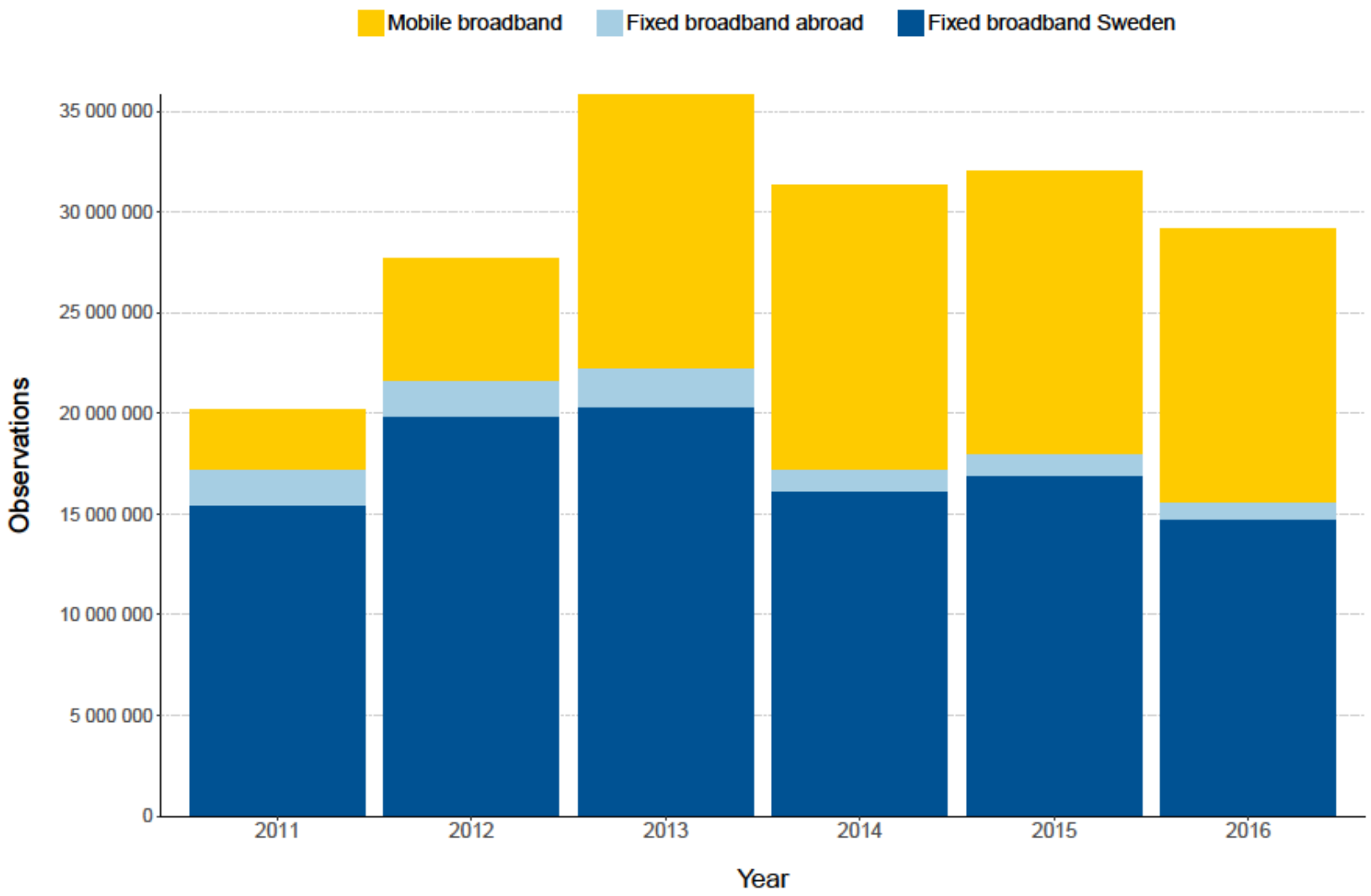
Index

Non-probabilistic samples



Index

Observations



Bredbandskollen records

USED

Unique ID | Client | Date/time | Country | Region

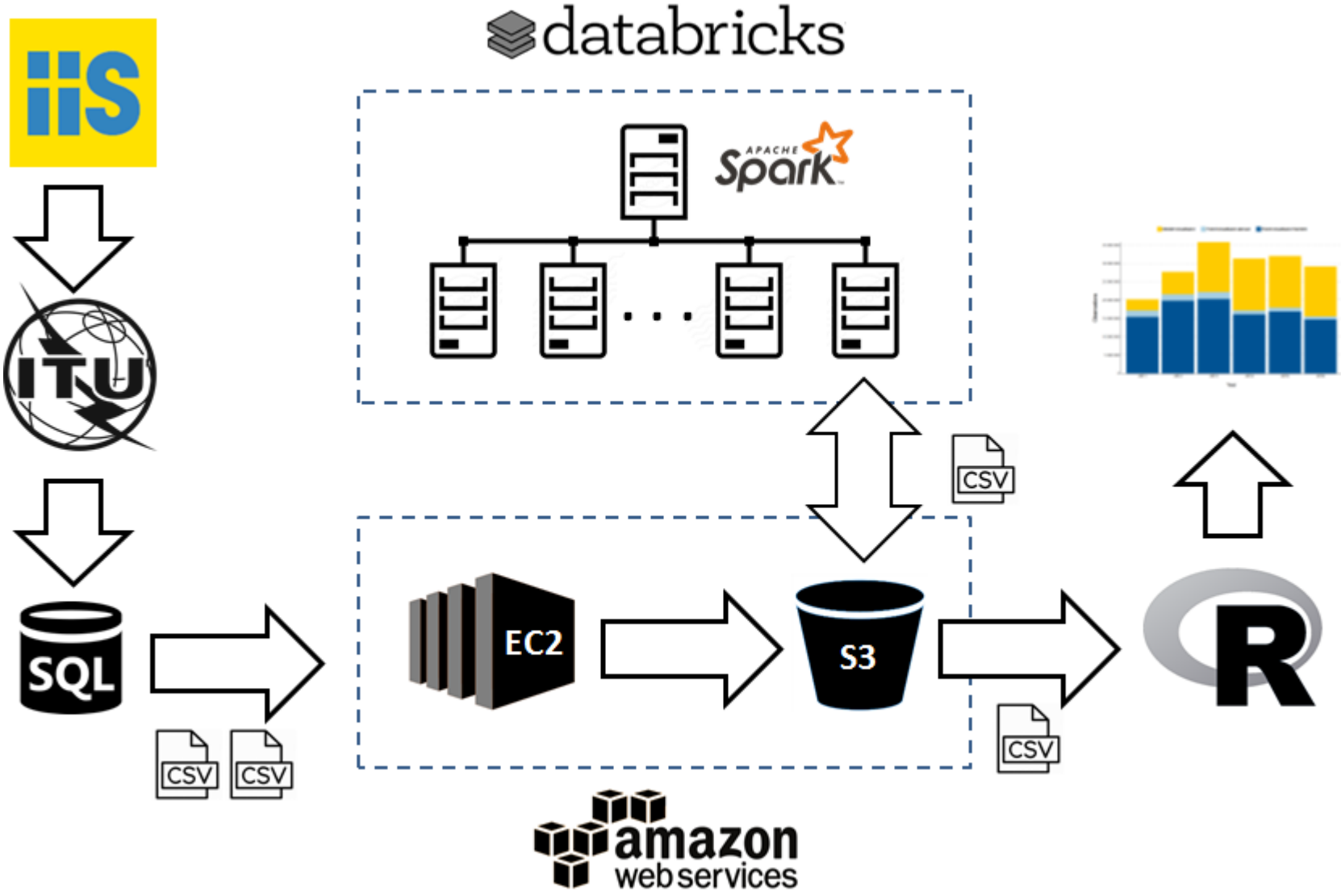
Download Speed | Upload Speed | Network type

Operator

NOT USED

Latency | Municipality

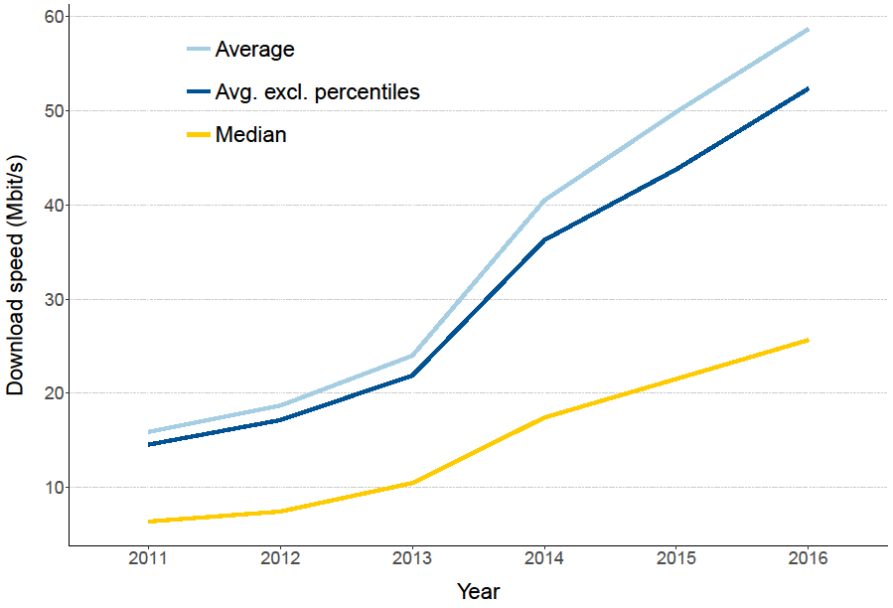
Data processing

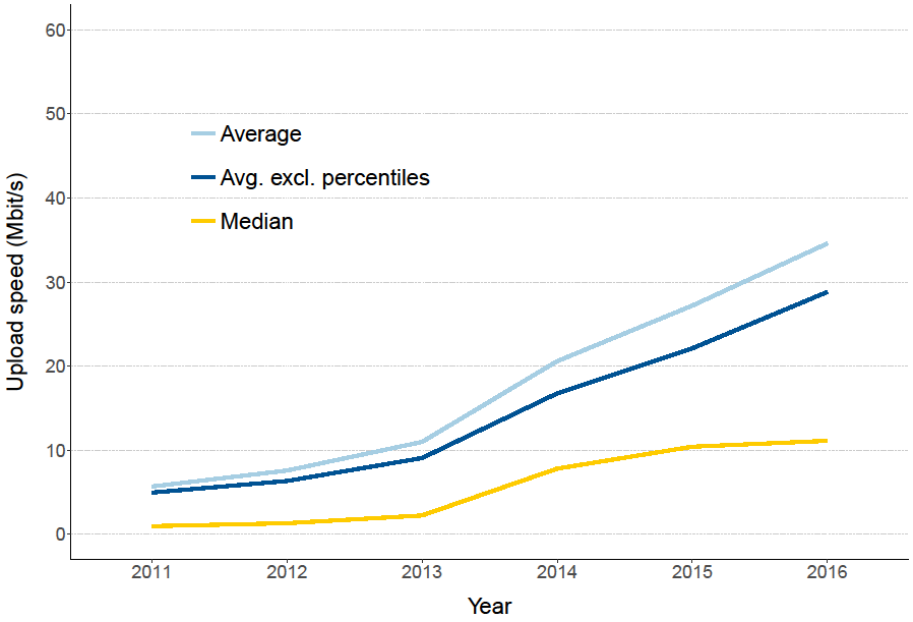


Index

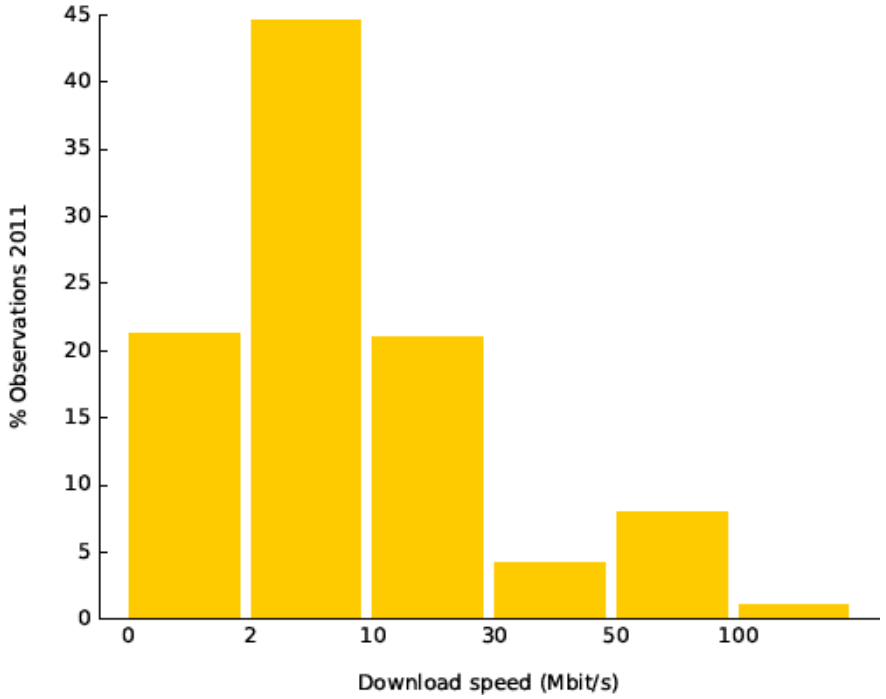
- Background & research question
- Methodology review
  - Broadband speed measurement platforms
  - Statistical inference from non-probabilistic samples
- Data set and environment
- **Results**
- Conclusions

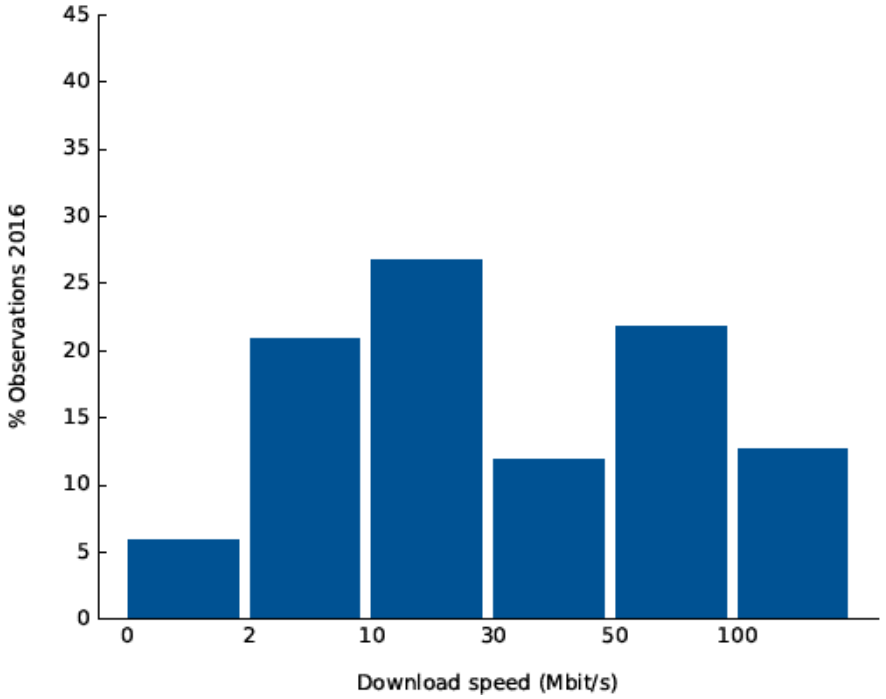
Aggregate speeds



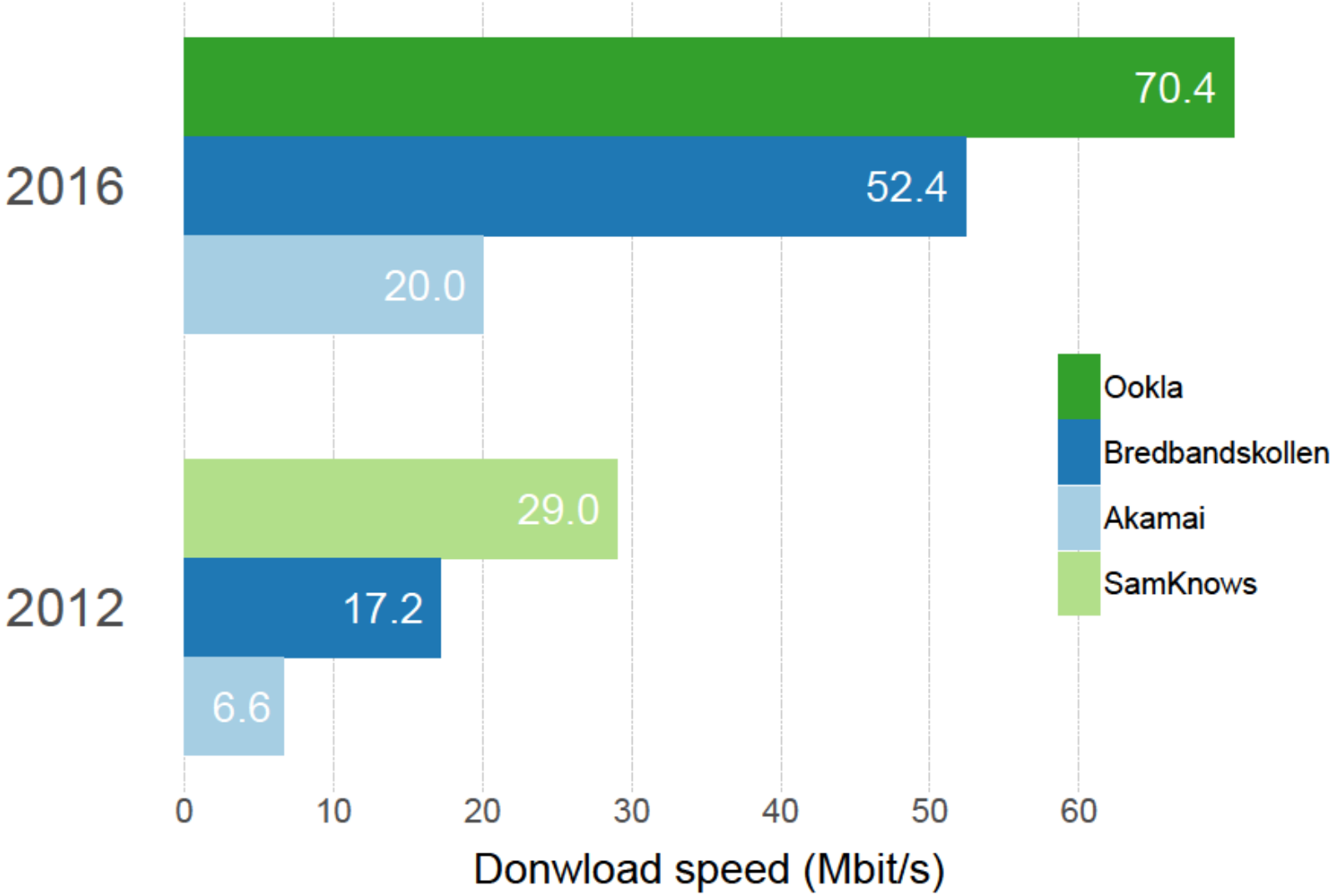


Histogram speeds

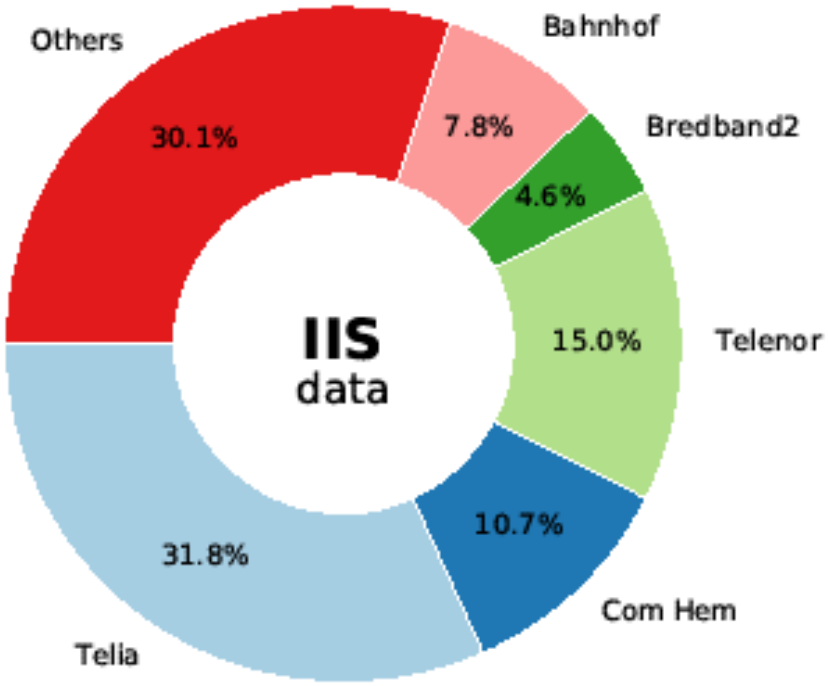


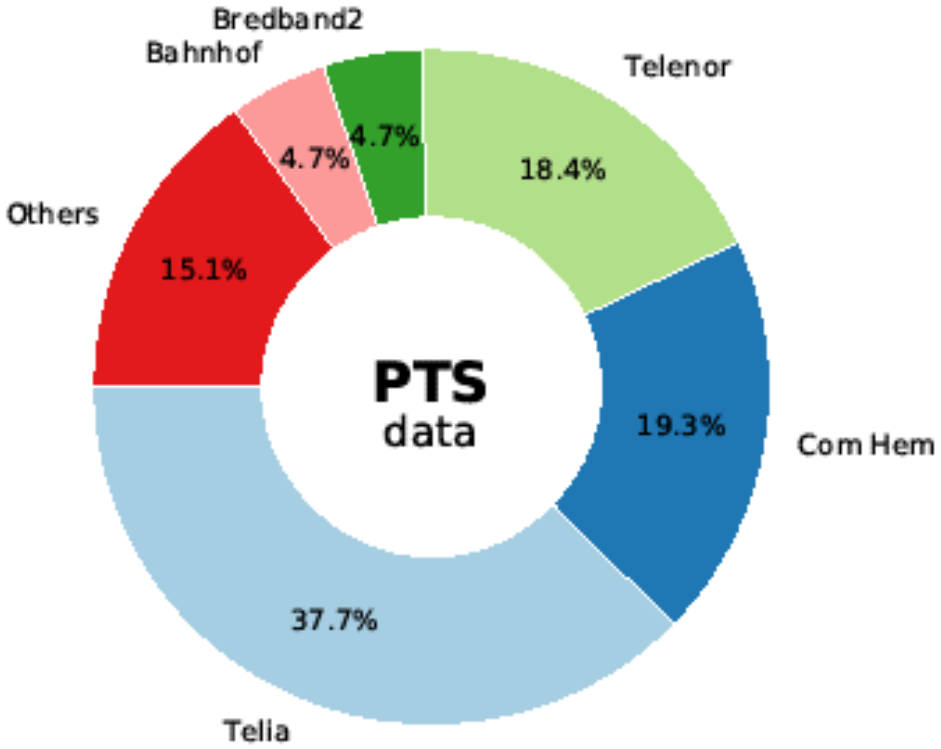


Comparison speed tests

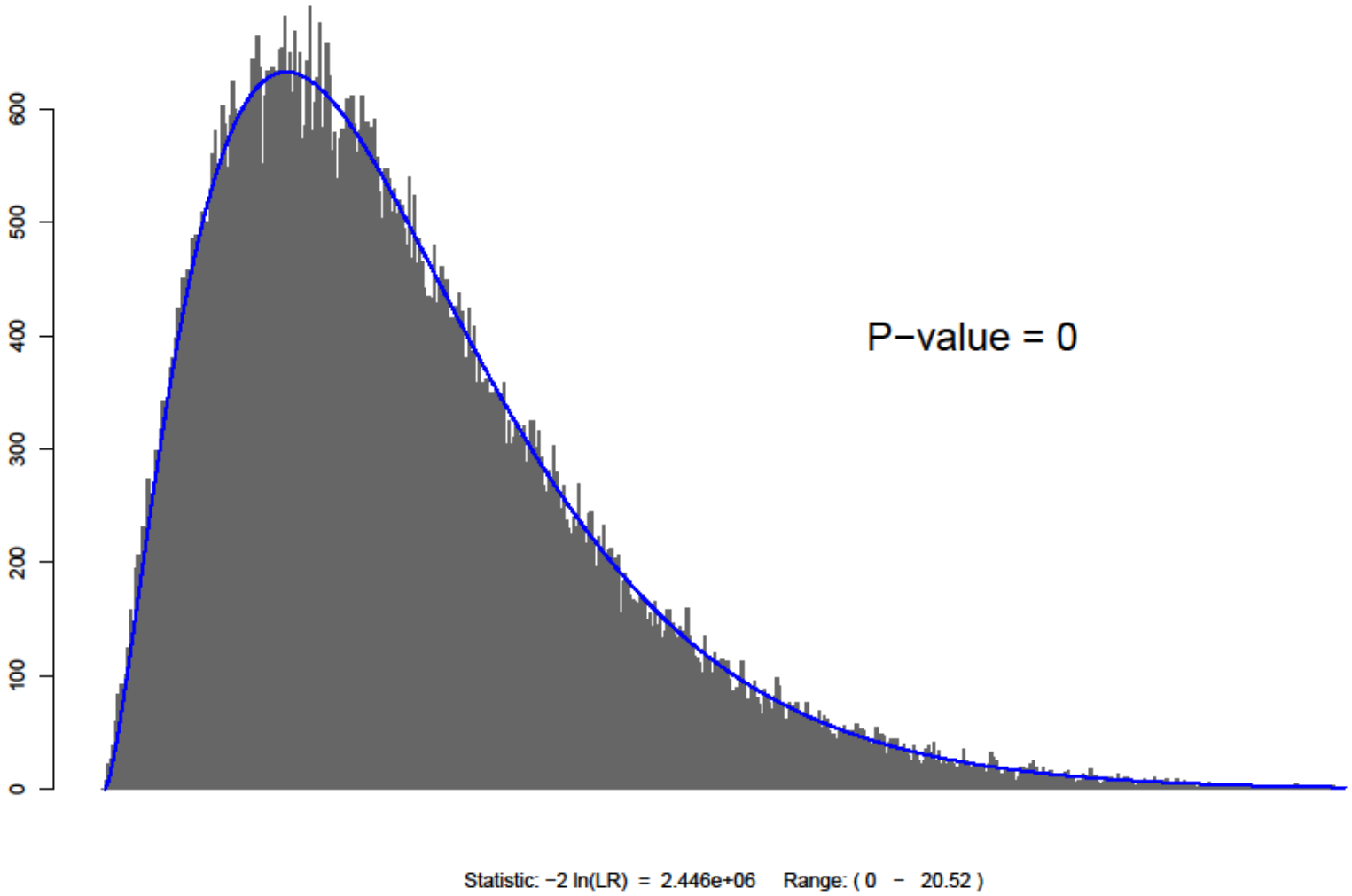


Observations per operator

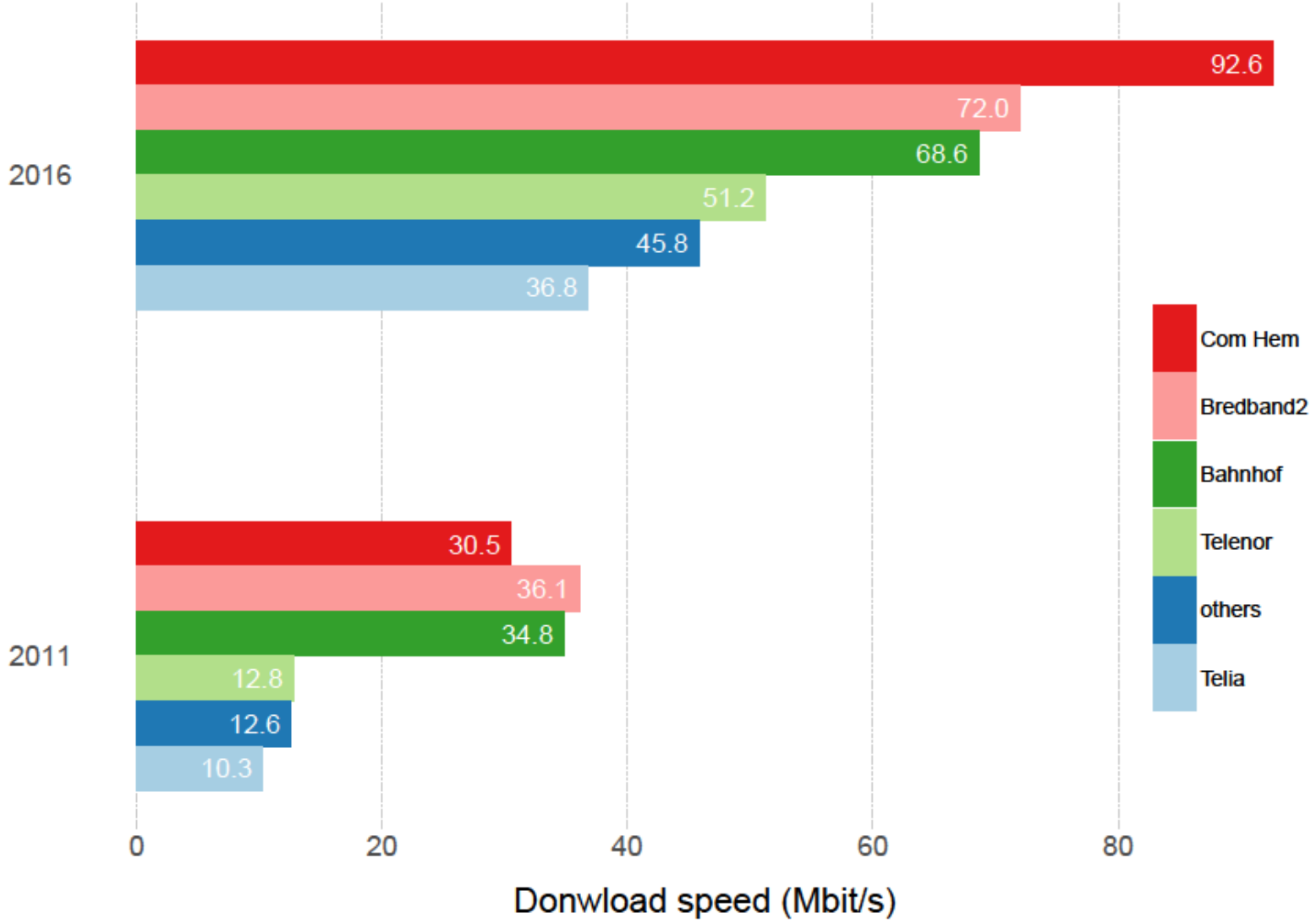




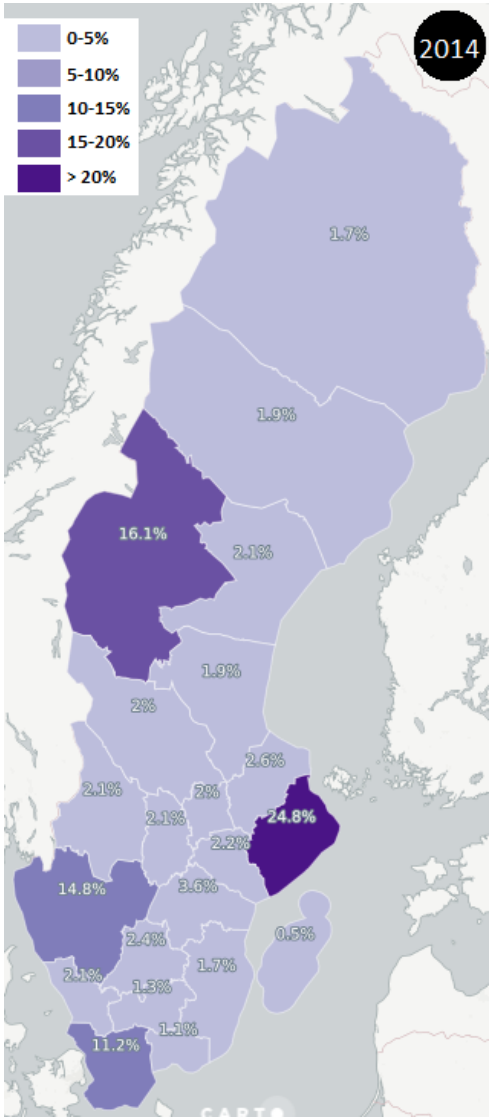
Selection bias

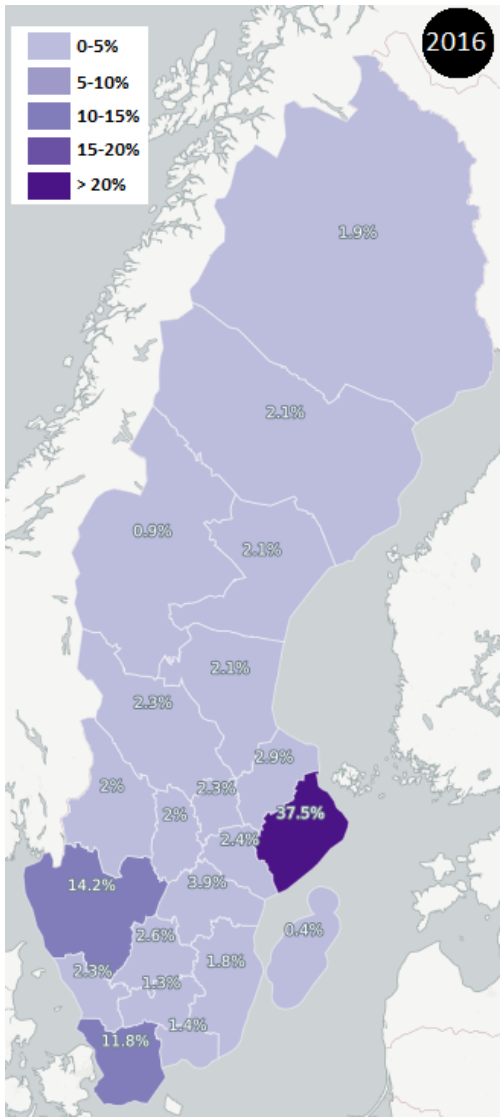


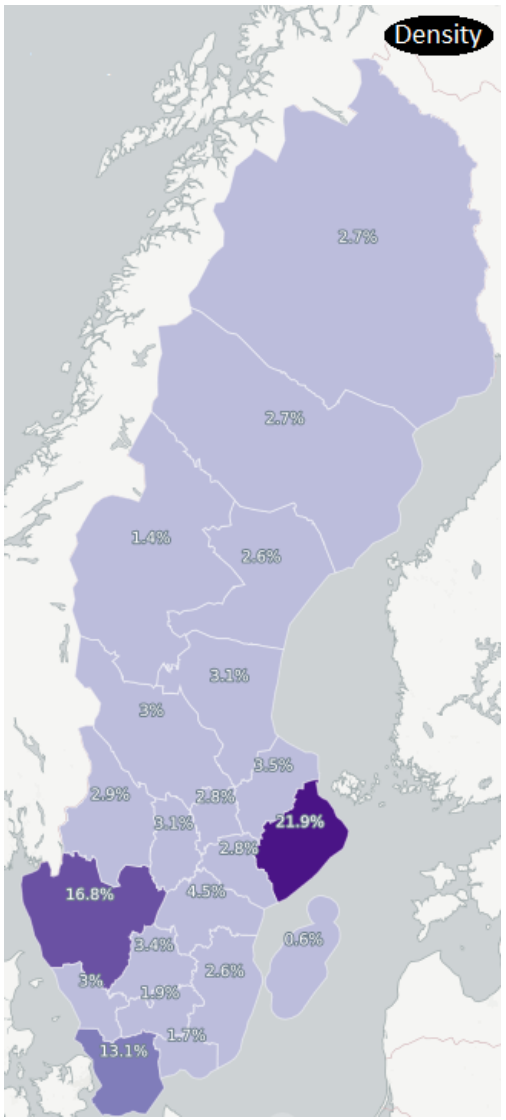
Speed by operator



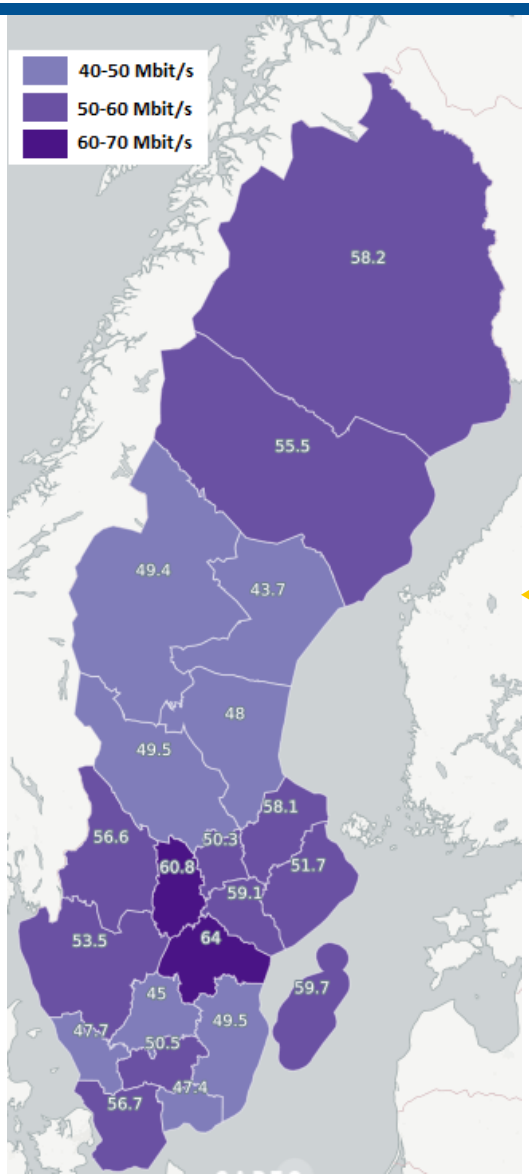
Observations per region





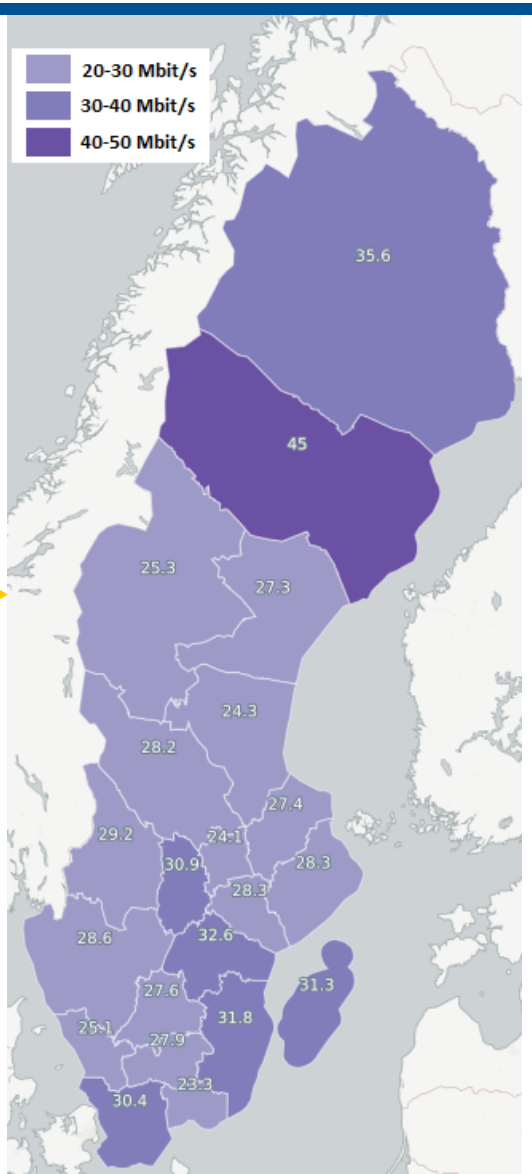


Speeds per region



Download

Upload



Index

Conclusions

Findings

- Selection bias
- Unstable sample composition
- In-house congestion (hypothesis)

Way forward

- + data: unique users, by technology
- + analysis: time, link socio-econ covariates

List of references

1. TNS Opinion & Social (2014). Special Eurobarometer 414 “E-Communications and Telecom Single Market Household Survey”. Technical report, European Commission.

2. Telecom Regulatory Authority of India (2006). Regulation quality of service standards for broadband services.

3. Davies, R. (2016). Broadband as a universal service. Technical report, European Parliamentary Research Service. PE 581.977.

4. Sundaresan, S., Feamster, N., and Teixeira, R. (2016). Home Network or Access Link? Locating Last-Mile Downstream Throughput Bottlenecks. In PAM 2016 - Passive and Active Measurement Conference, pages 111–123, Heraklion, Greece.

5. M-Lab Research Team and others (2014). ISP interconnection and its impact on consumer Internet performance – a measurement lab consortium technical report.

6. Rood, H., Yoshikawa, D., Wevers, R., Post, R., Tetteroo, A.-J., Kuipers, A., and Schurmann, R. (2012). Usability of broadband performance measurements for statistical surveys. In 2012 TRPC.

7. Couper, M. (2013). Is the sky falling? new technology, changing media, and the future of surveys. Survey Research Methods, 7(3):145–156.

8. Zagheni, E. and Weber, I. (2015). Demographic research with non-representative internet data. International Journal of Manpower, 36(1):13–25.

Thank you