

MASTER PROJECT

Measuring real broadband speeds using
crowdsourcing data from the Internet Foundation

Author: Ivan Vallejo Vall

Tutor: Iñigo Herguera García

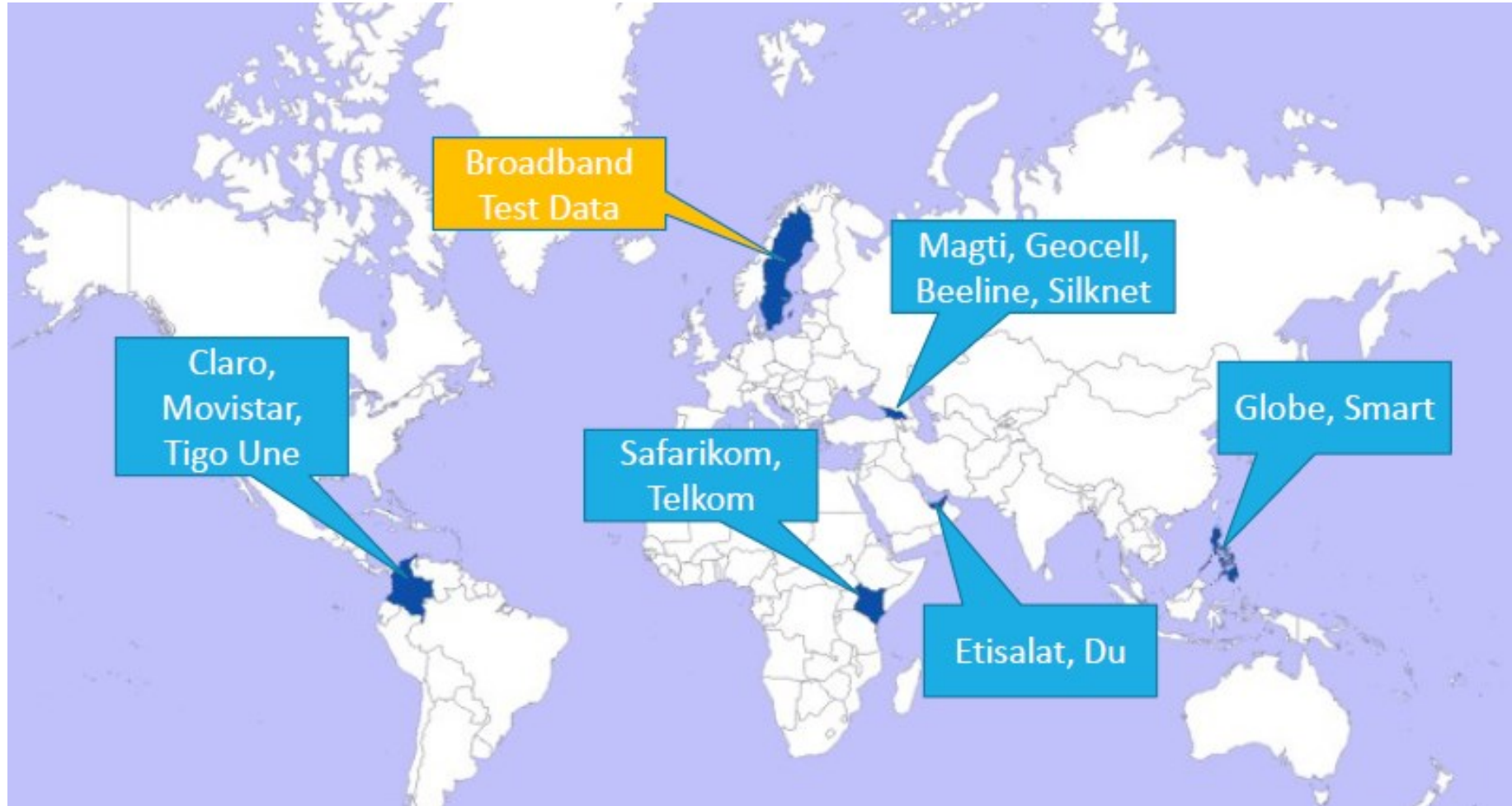
Data Science Program 2016/2017

Index

2

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

Big Data for Measuring the Information Society



Bredbandskollen

4

A screenshot of the Bredbandskollen website interface. The header is blue with the title 'Bredbandskollen' in white. Below the header is a navigation bar with icons and labels for 'HOME', 'SELF HELP', 'SPEEDMAP', 'REPORTS', 'HELP', and a lightbulb icon. The main content area has a blue background with a large clock icon on the right. The text on the left says 'Bredbandskollen provides a simple way for you to measure your internet speed and get help on improving it.' Below this is a green button that says 'START TEST!'. At the bottom, there are two sections: 'Bredbandskollen on your phone' with a smartphone icon and text about downloading the app for iOS or Android, and 'Follow Bredbandskollen' with a Facebook icon and text about following them on Facebook and Twitter.

Bredbandskollen

HOME SELF HELP SPEEDMAP REPORTS HELP

Bredbandskollen provides a simple way for you to measure your internet speed and get help on improving it.

START TEST!

Bredbandskollen on your phone
Fast internet on your phone? Get our app for [IOS](#) or [Android](#)

Follow Bredbandskollen
If you also think Bredbandskollen is great, you can follow us on [Facebook](#) and [Twitter](#).

Relevance of Internet speed

5

□ Consumers

- 2nd most important factor
- advertised \neq real speed + congestion [1]

□ Regulators

- Quality of service (QoS) [2]





□ Policy makers

- Universal Service Funds, broadband plans [3]

Research question

6

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

- ❑ Advertised speeds 
- ❑ Hardware-based measurements   
- ❑ Software-based crowdsourcing?

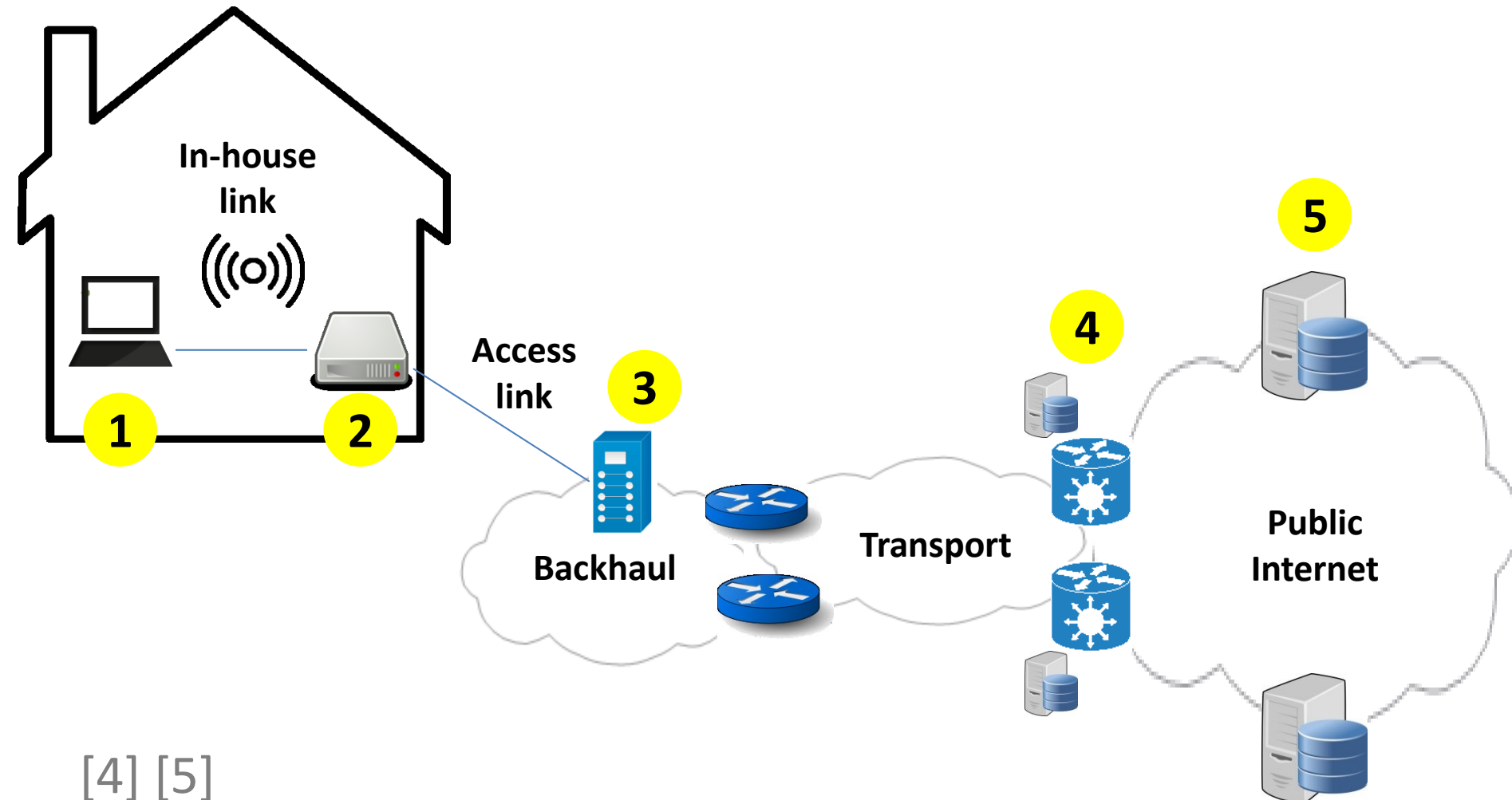
Index

7

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

From where to where?

8



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

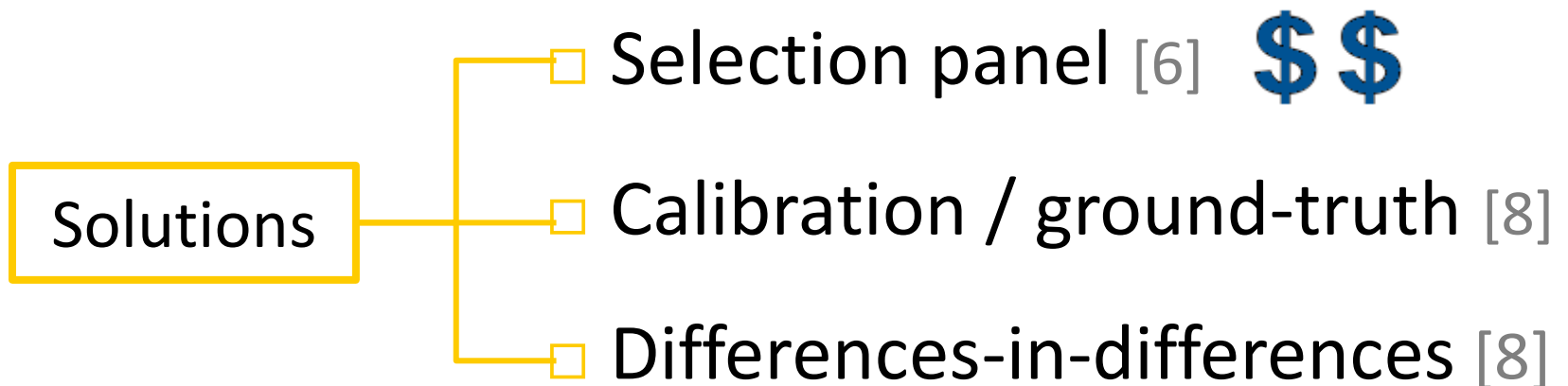
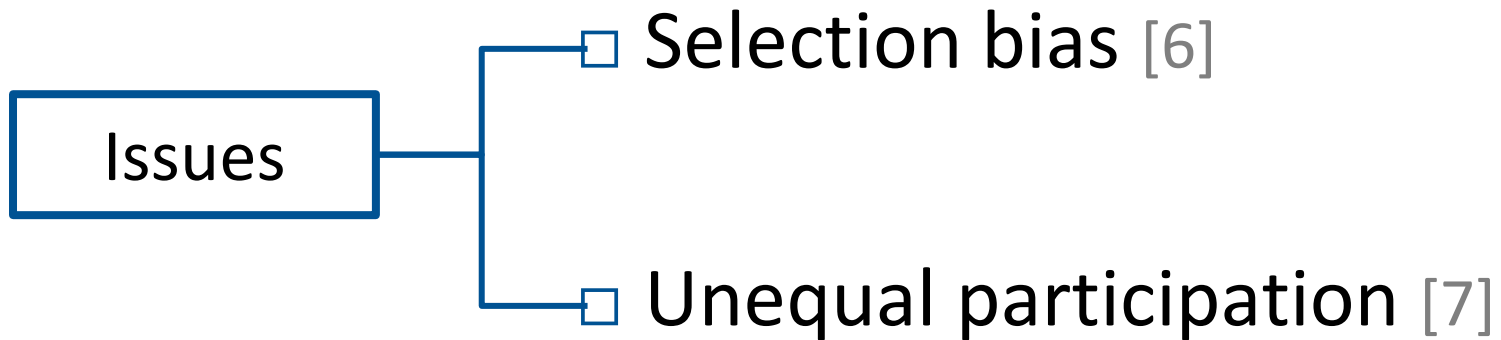
Index

10

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

Non-probabilistic samples

11



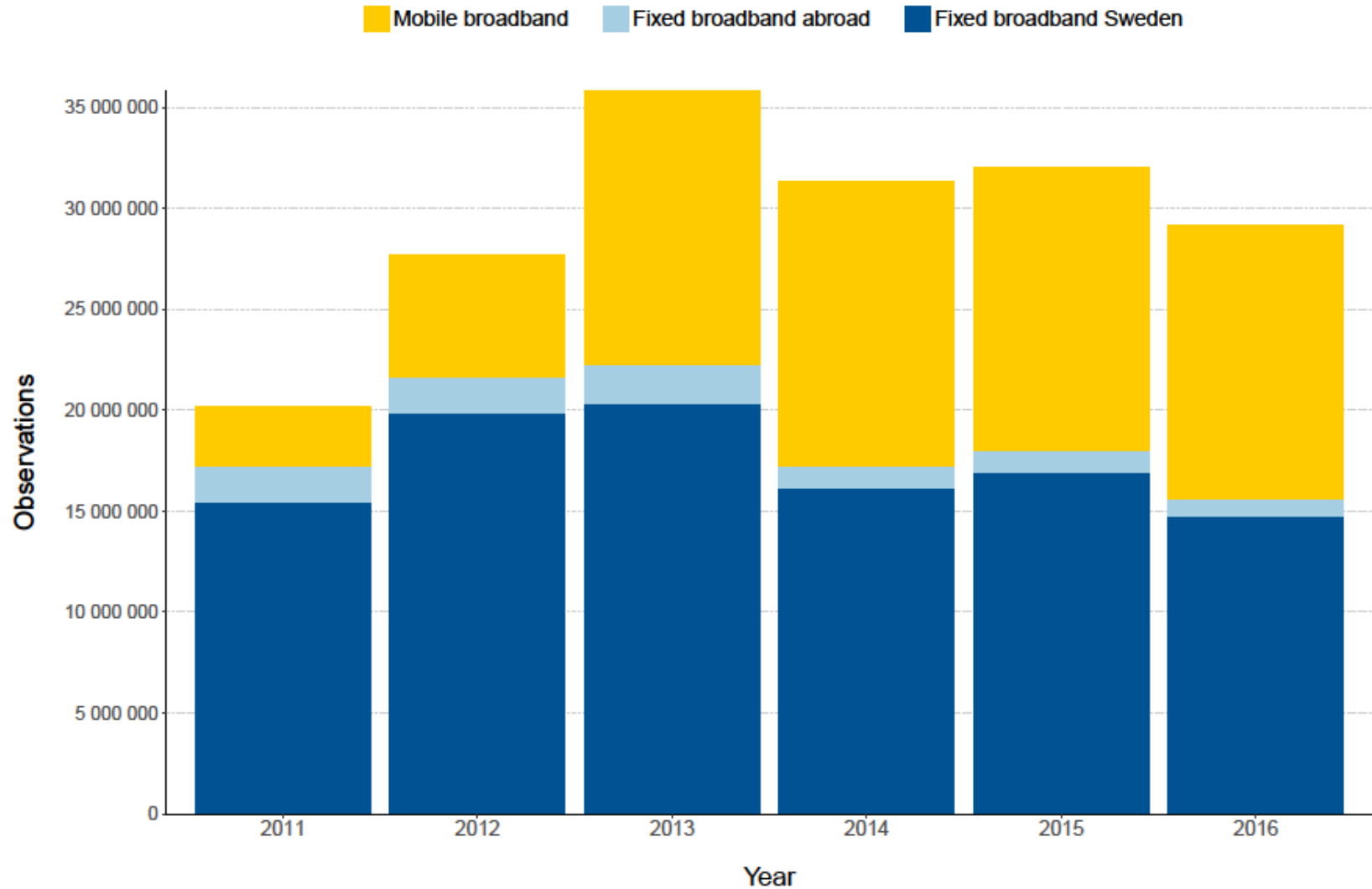
Index

12

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

Observations

13



Bredbandskollen records

14

USED

Unique ID

Client

Date/time

Country

Region

Download Speed

Upload Speed

Network type

Operator

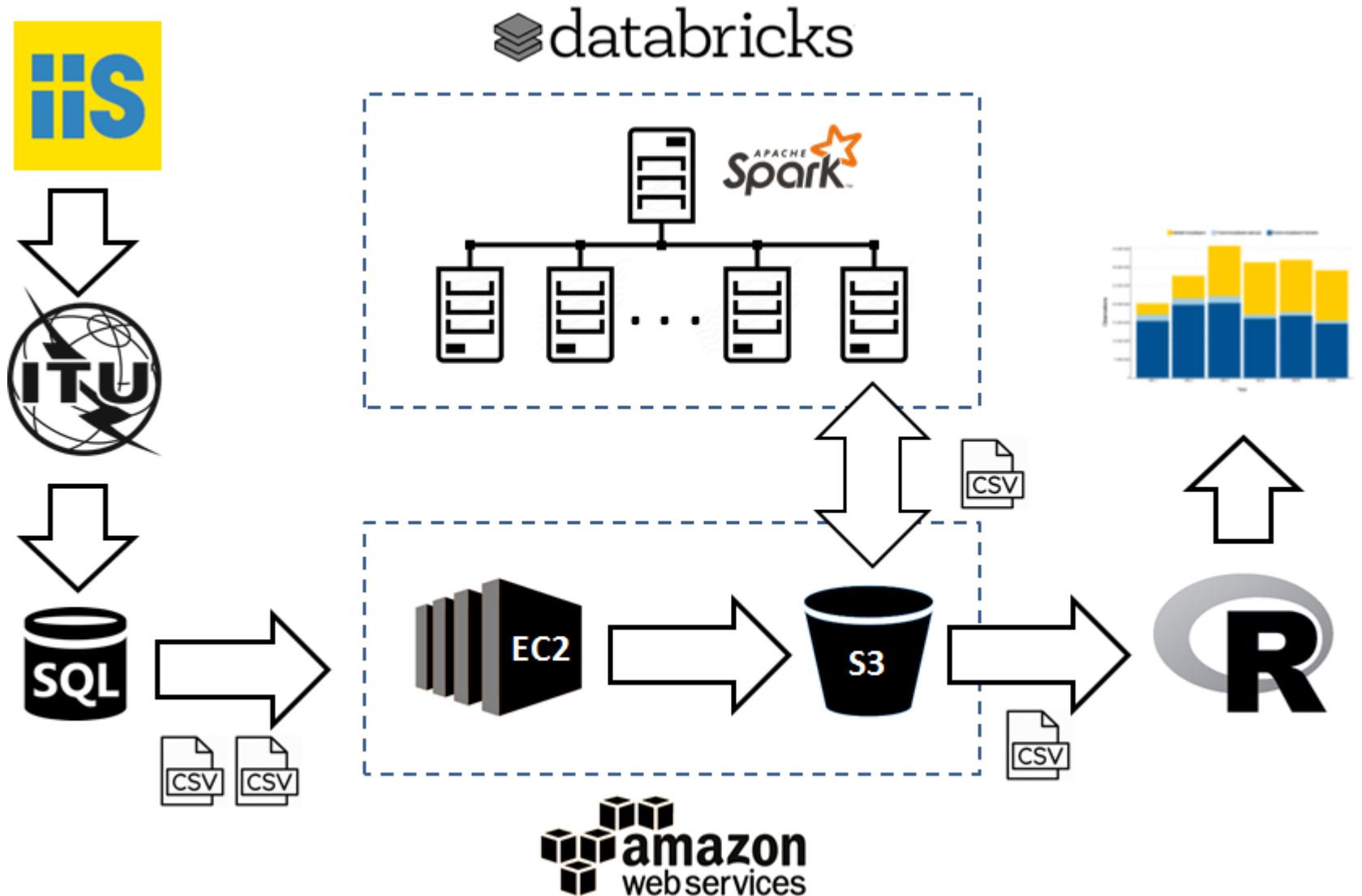
NOT USED

Latency

Municipality

Data processing

15



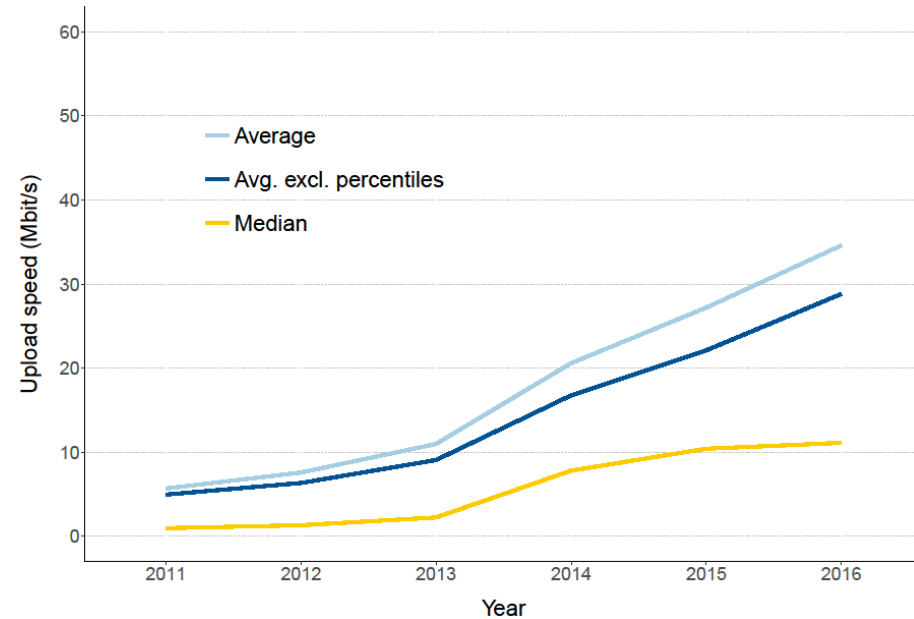
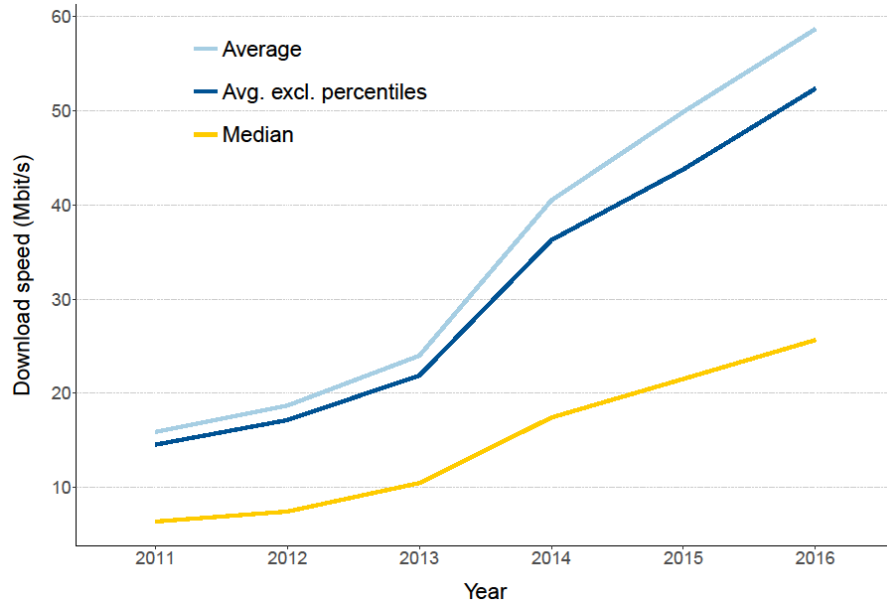
Index

16

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

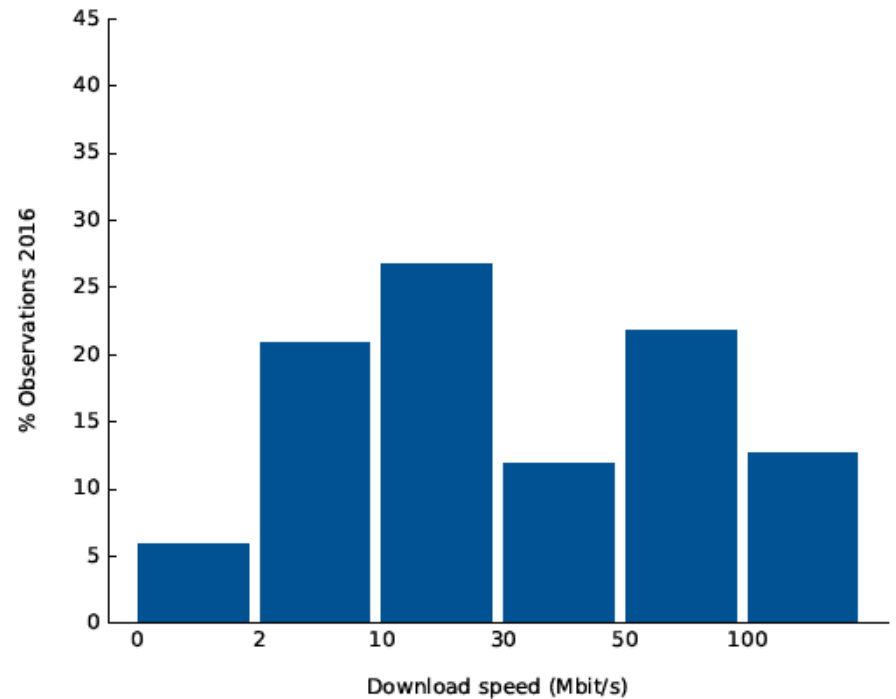
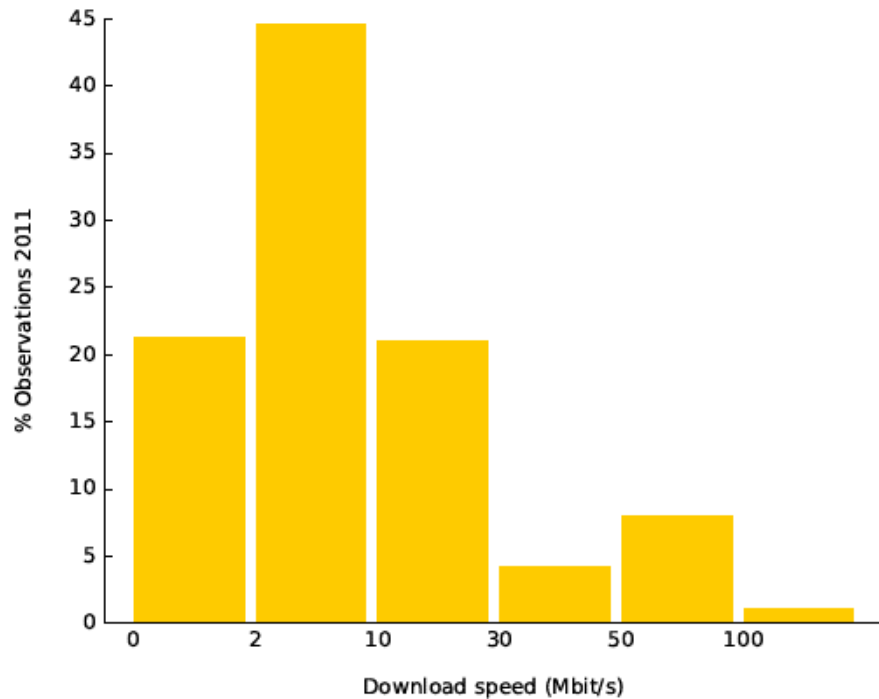
Aggregate speeds

17



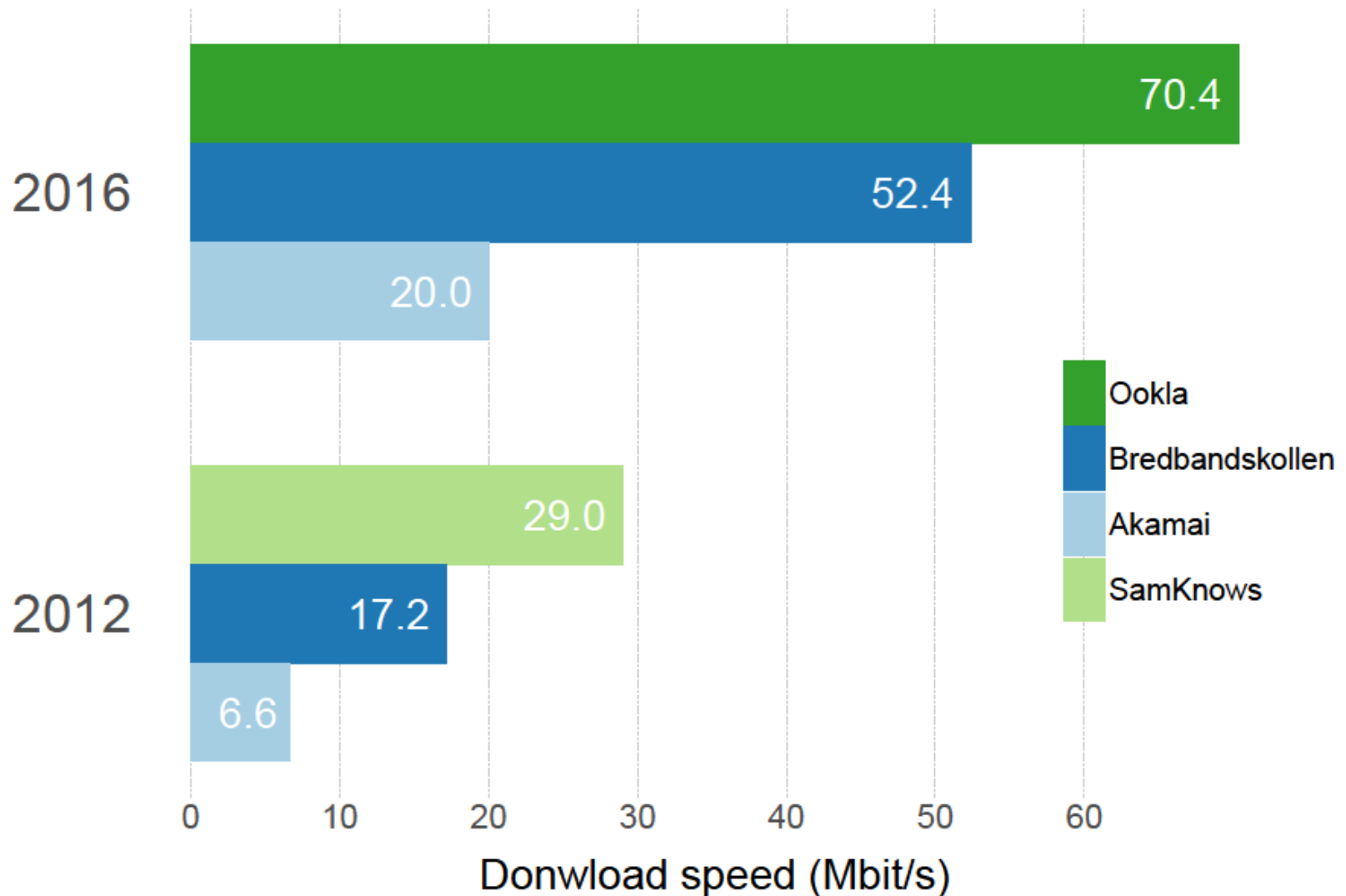
Histogram speeds

18



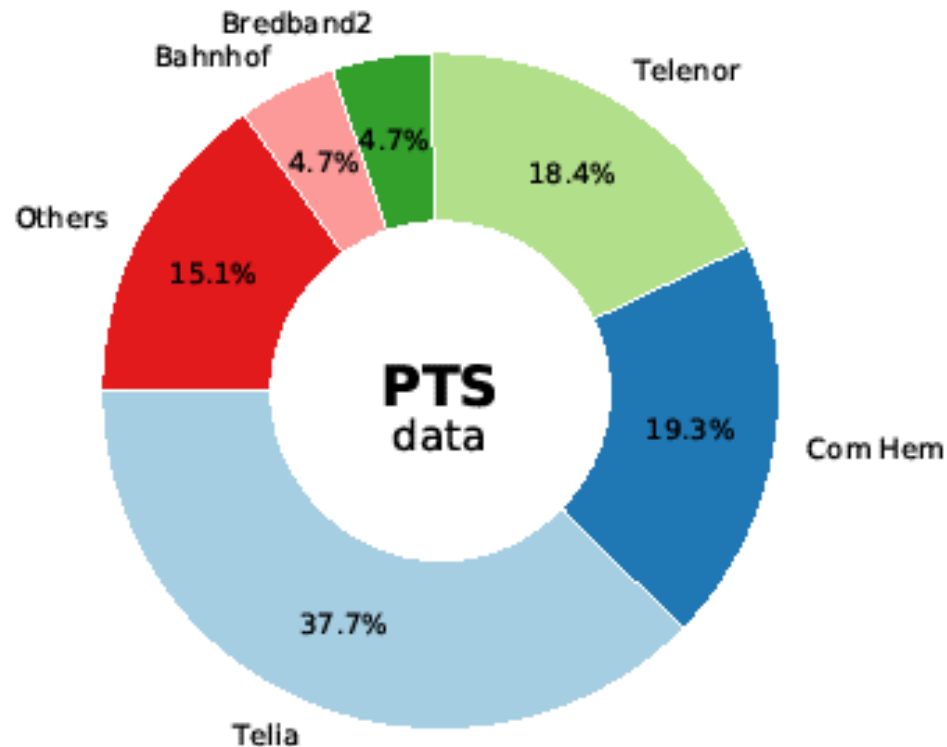
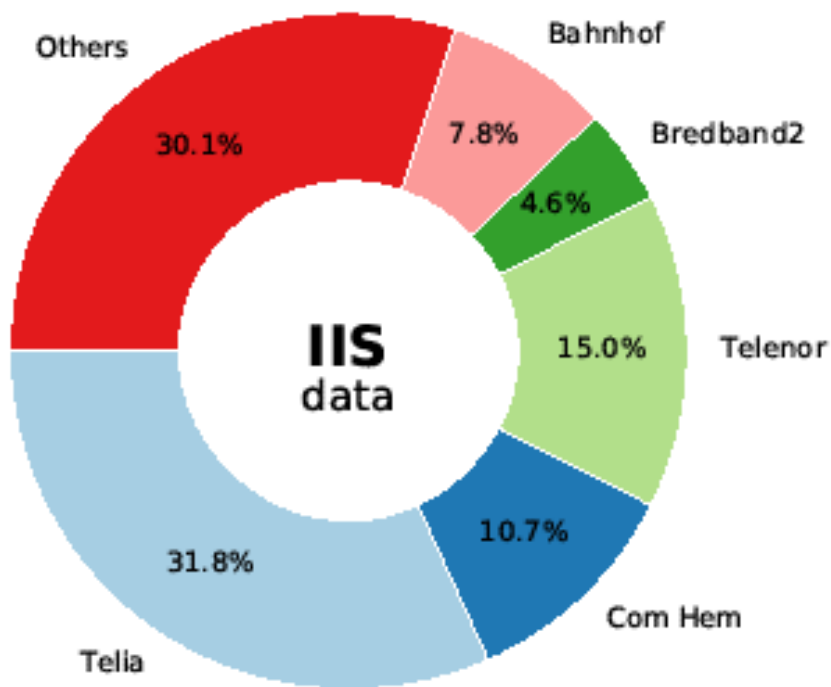
Comparison speed tests

19



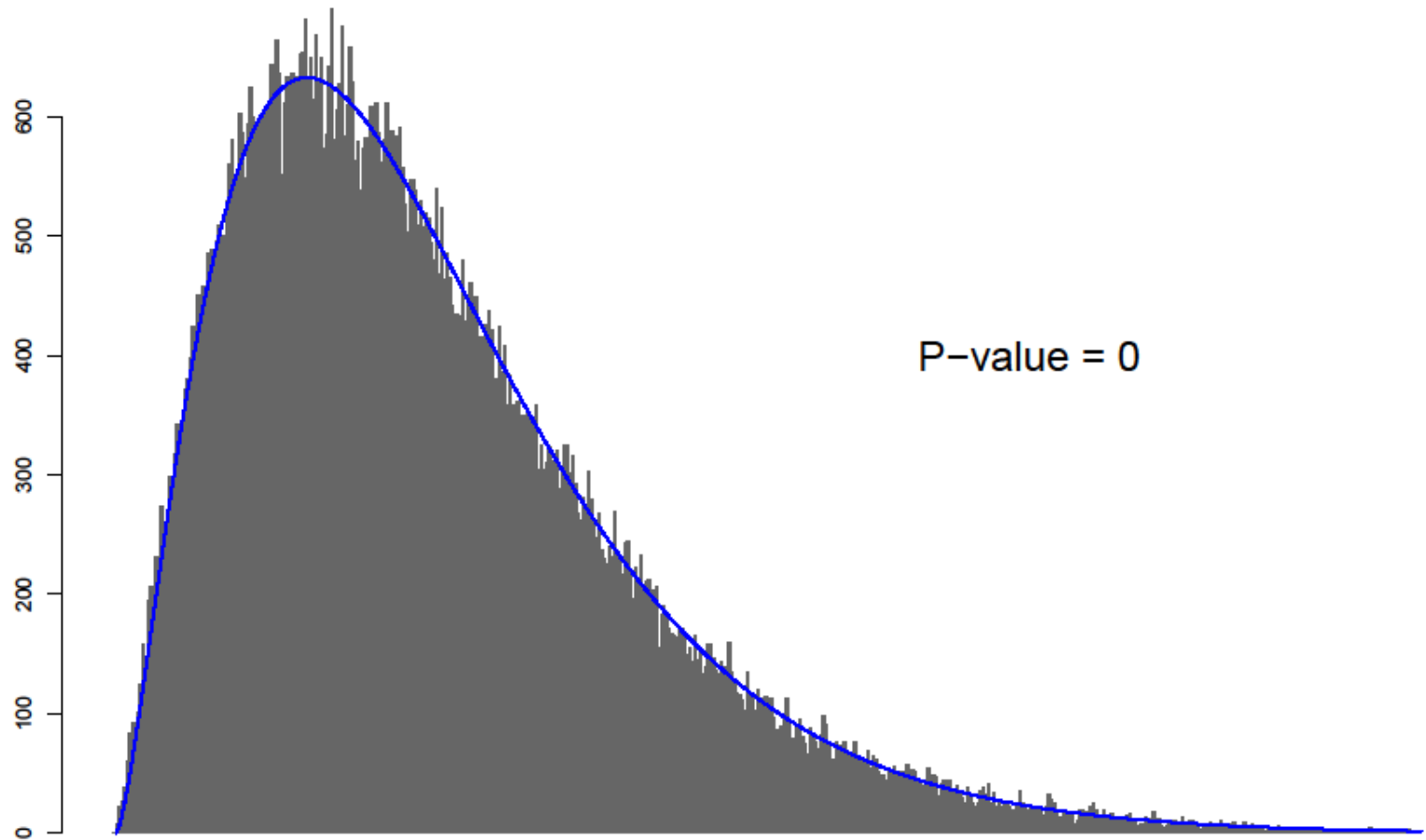
Observations per operator

20



Selection bias

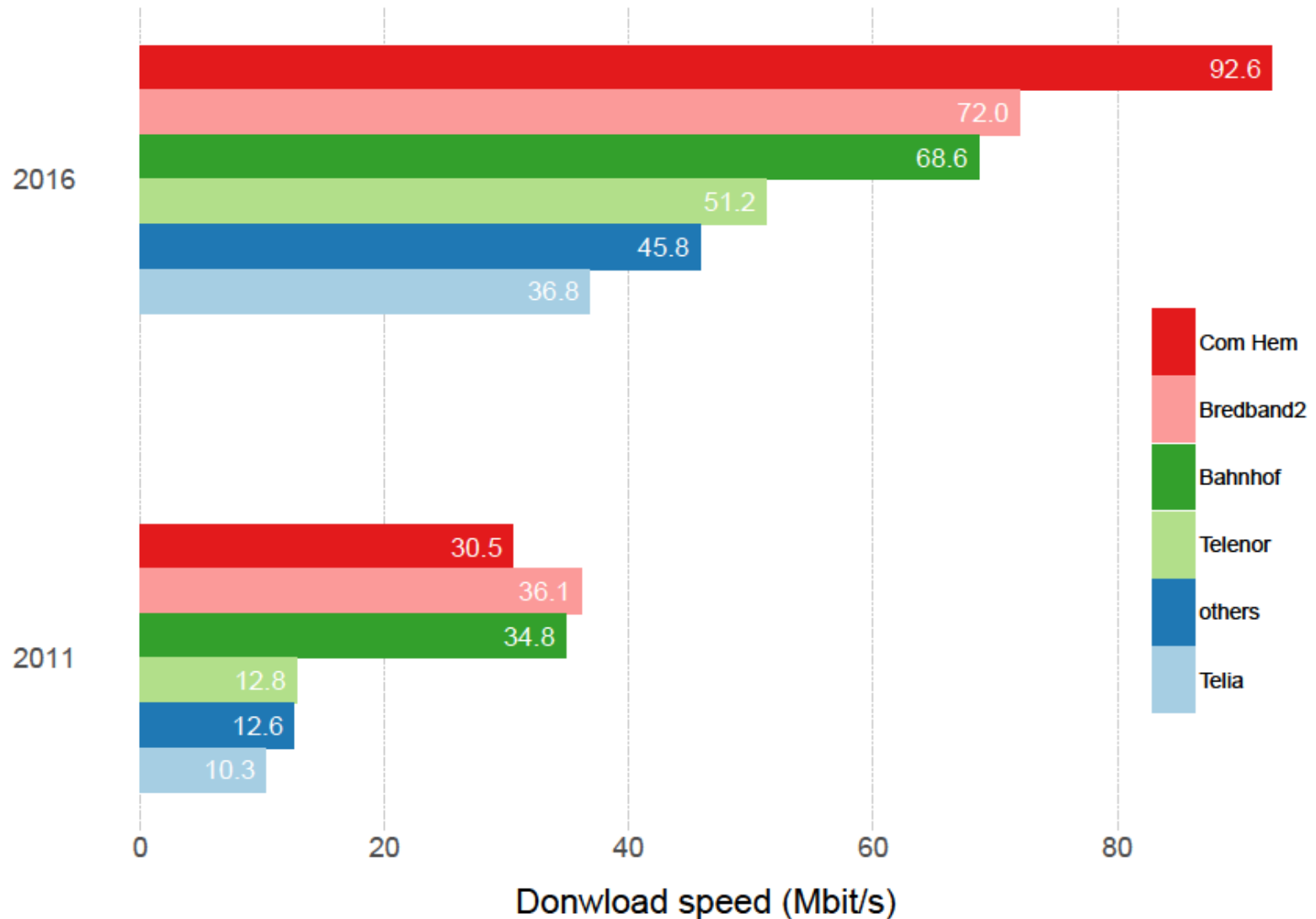
21



Statistic: $-2 \ln(LR) = 2.446e+06$ Range: (0 - 20.52)

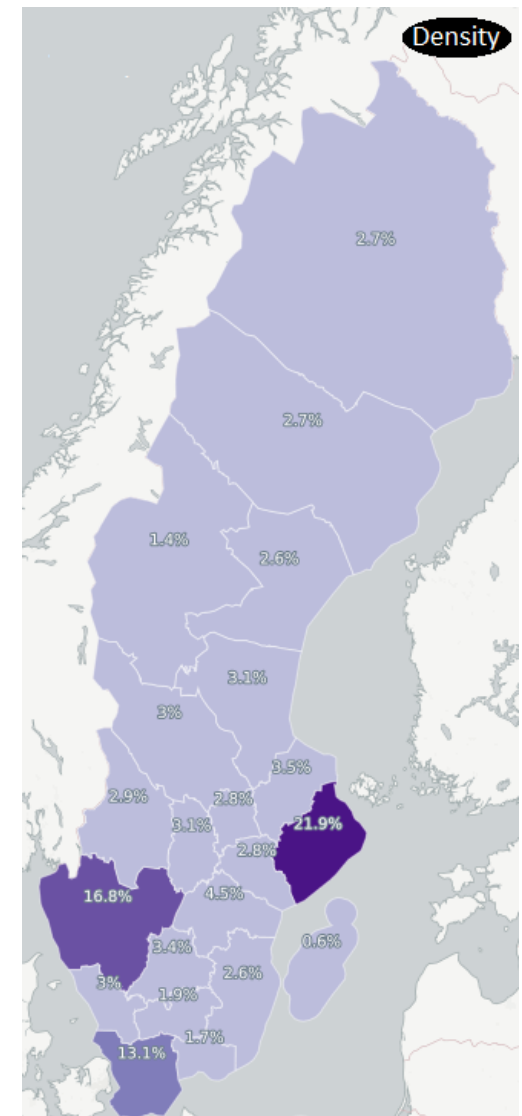
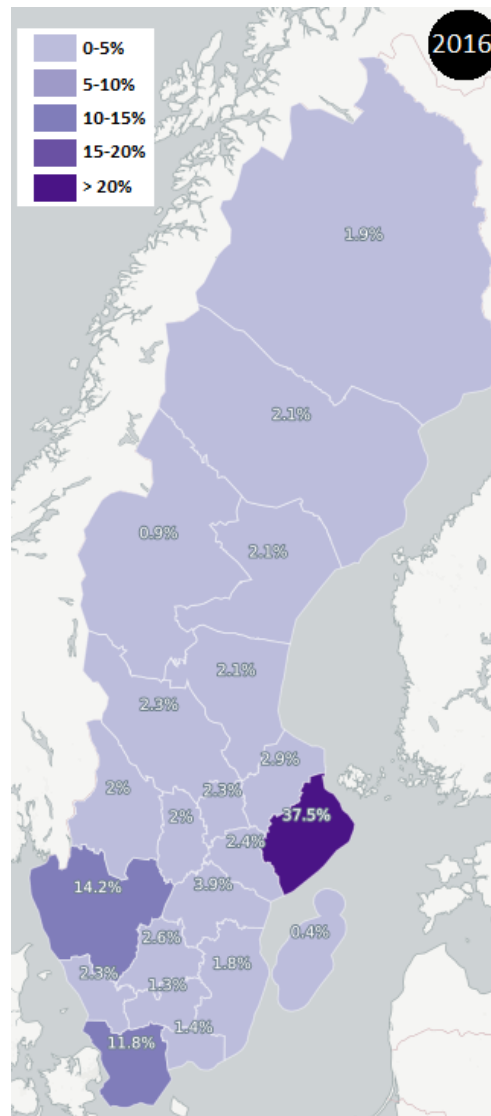
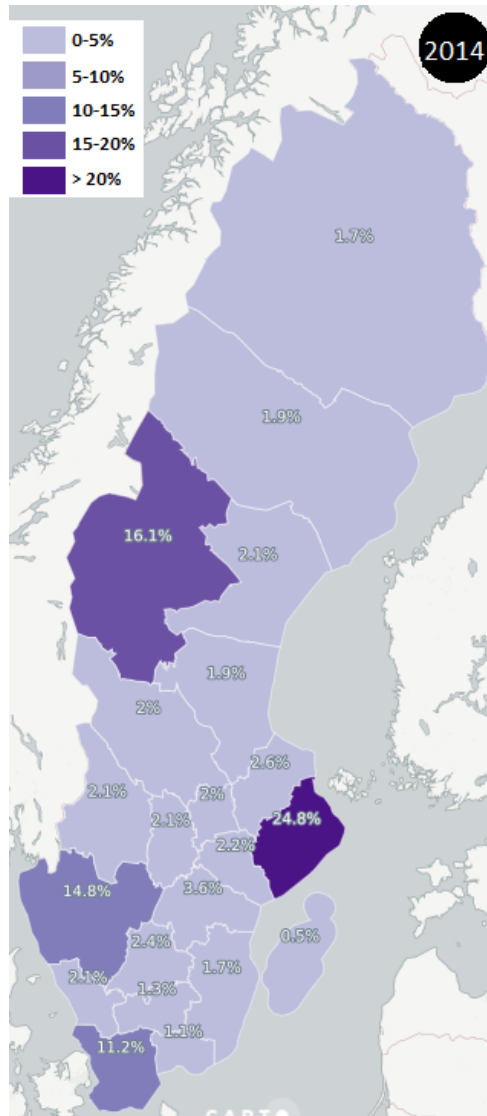
Speed by operator

22



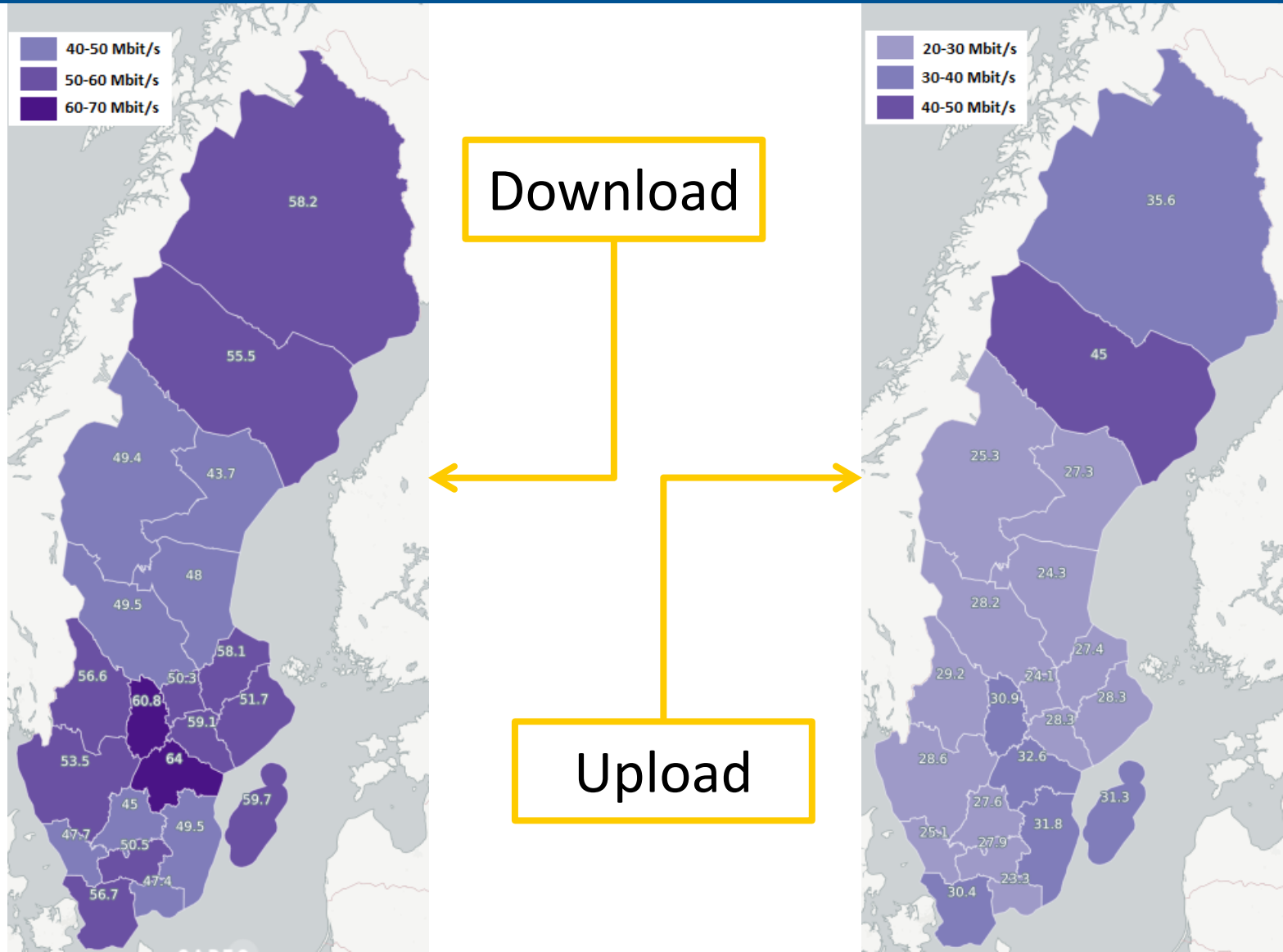
Observations per region

23



Speeds per region

24



Index

25

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

Conclusions

26

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

27

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