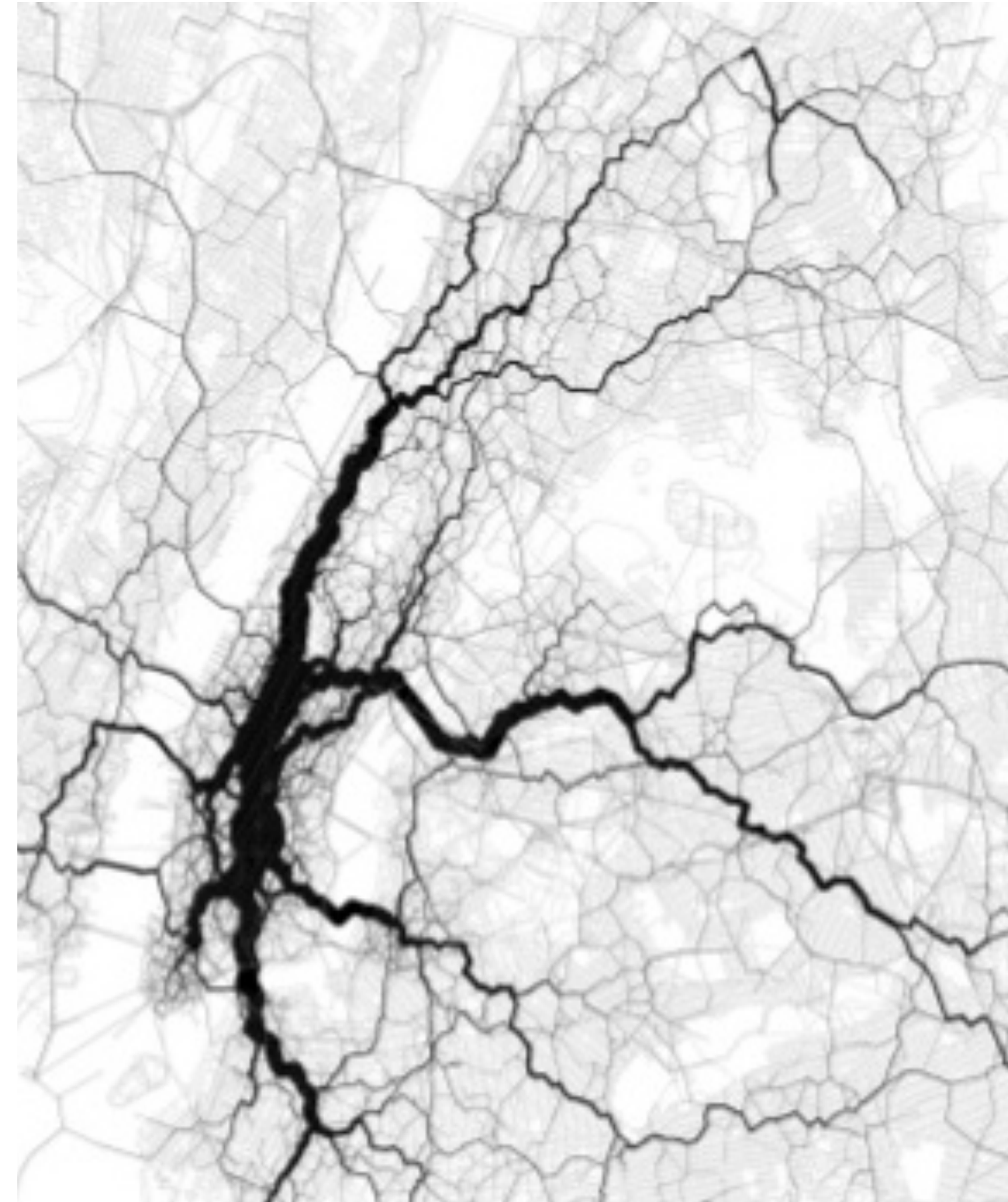


IT UNIVERSITY OF COPENHAGEN

Reflections in Data Science

BSREDAS1KU-20201

2020-02-18



Exercise 2 - Plan

- More details about Fermi estimation method.
- Two Fermi estimations problems (10 minutes each).
- A data visualisation problem (20 minutes).

When does Fermi estimate fails?

- Badly overestimate or underestimate
- Model the problem incorrectly
- You might be facing a non-linear problem.

Recommendations

- Start with estimations you can check later.
- Don't look the answer before centring your estimations, though!
- You can check pieces of the problem.

Fermi problem 1

- How many bald persons we have in the world?

Fermi problem 1

- World's population?
- Portion of the population that might go bald?

Fermi problem 2

- How many bicycles are sold in Denmark every year?

Fermi problem 2

- Danish population?
- Portion of population interested in bicycles?
- Percentage of new buys?

Fermi problem 2

- How many bicycles are sold in Denmark every year?

Answer: Approximately half million bicycles are sold every year in Denmark

Source: <http://www.cycling-embassy.dk/wp-content/uploads/2010/03/Bicycle-statistics-from-Denmark.pdf>

Visualisation Study

- A New study says that exercise is associated to increased levels of cholesterol.

www.mybinder.org

<https://github.com/tocunha/reflectionsdatascience>

Simpson Paradox

- Trend appears in several different groups, but disappears or reverses when these groups are combined.

Simpson Paradox

https://www.youtube.com/watch?v=E_ME4P9fQbo