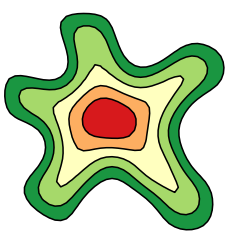
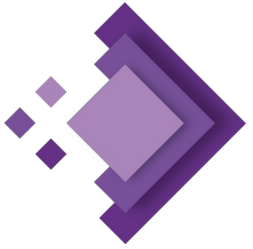
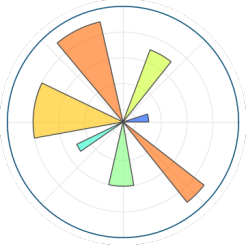
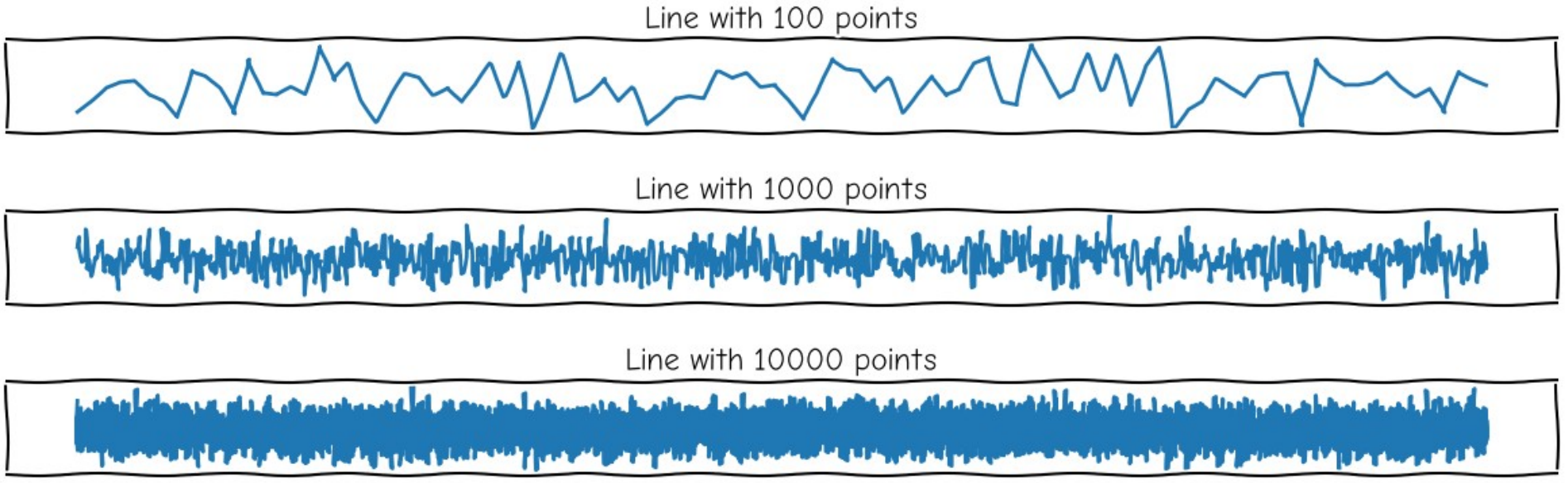
Visualising large datasets: everyone in the UK Census

Ian Thomas
Caerphilly
Anaconda

ANACONDA.

The problem of visualising large datasets



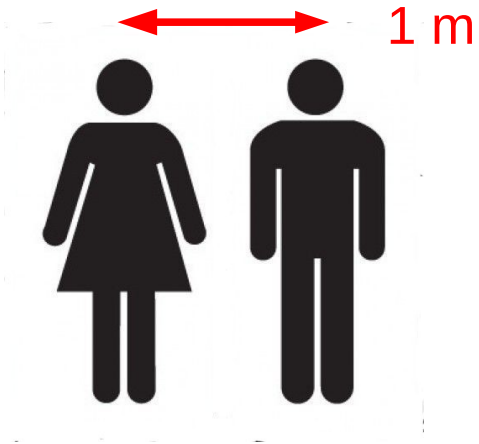
This is called overplotting

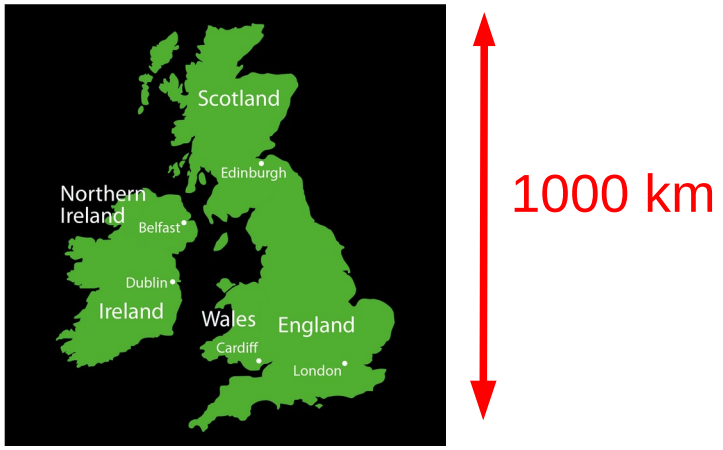
ANACONDA.

Real data: UK Census 2011

- Relevant to everyone in the UK
- 63 million people
- (x, y) locations and some categorical data
- Want to know where they are, and something about them, visually
- Source: Office for National Statistics

How many pixels required?





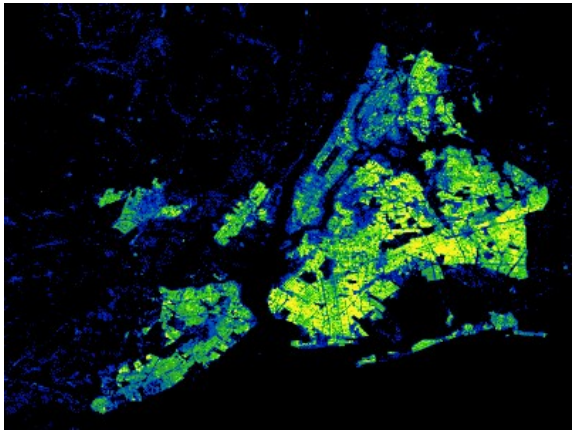
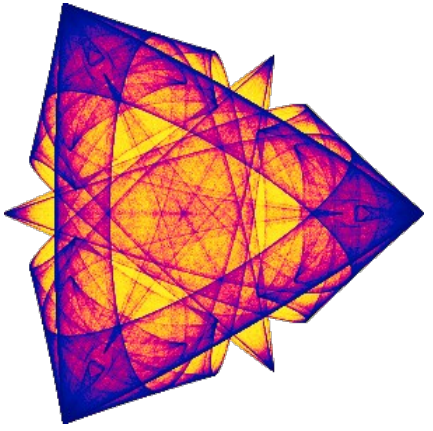
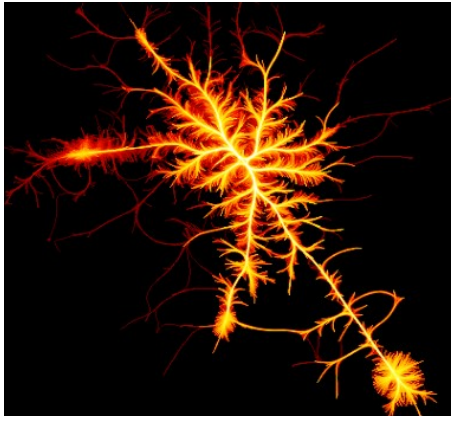
- Need ~1 million pixels to resolve everyone!
- Datashader invented to solve problems like this.

ANACONDA.

Show some code Ian

What else can Datashader do?

- Other reductions: https://datashader.org/api.html#reductions
- Lines, polygons, triangle meshes, etc

- Can use Dask or cupy for larger datasets and better performance

ANACONDA.

Summary

- Introduction to Datashader using UK Census data
- Count of people per pixel
- Nonlinear shading using a colourmap
- Counts by category
- hvPlot for interactive exploration

ANACONDA.

Contact information

Ian Thomas

GitHub repository

https://www.github.com/ianthomas23/pycon-uk-2022

ANACONDA.