

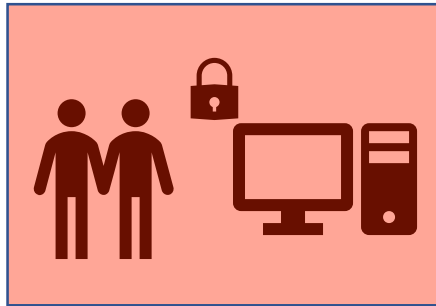
SDC Handbook

Simon Parker, Cancer Research UK

Richard Welpton, The Health Foundation

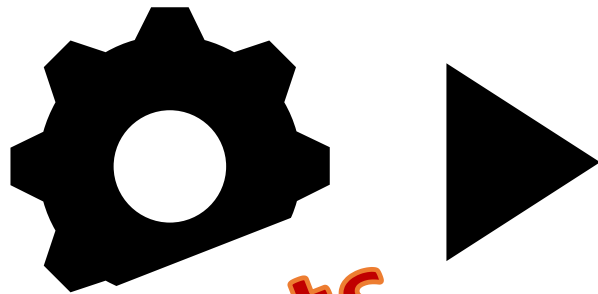


Introduction

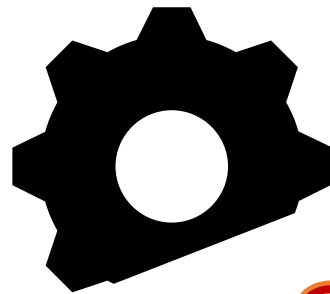
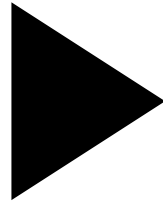


SDC

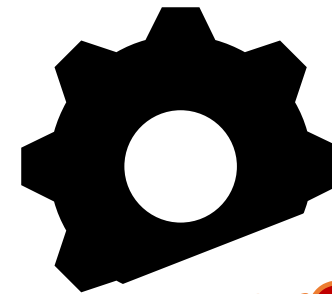
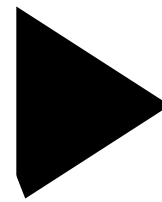
HERE



Inputs



Analysis



Outputs

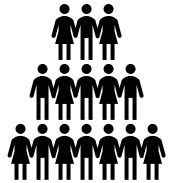
Previous guidance

Theoretical considerations and a **few practical resources**, e.g. ESSnet/DwB

SDC: **no longer confined to a small group of people**



Training is essential (and Felix Ritchie will discuss this later)



WANTED:

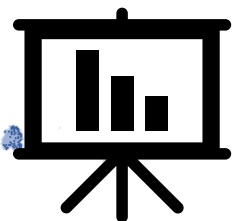
An easy to use set of guidance that somebody without expertise can pick up and apply



Working Group for Safe Data Access Professionals.....

Putting together a new SDC Handbook.....

Features



Range of common statistical outputs reviewed



Traffic light system



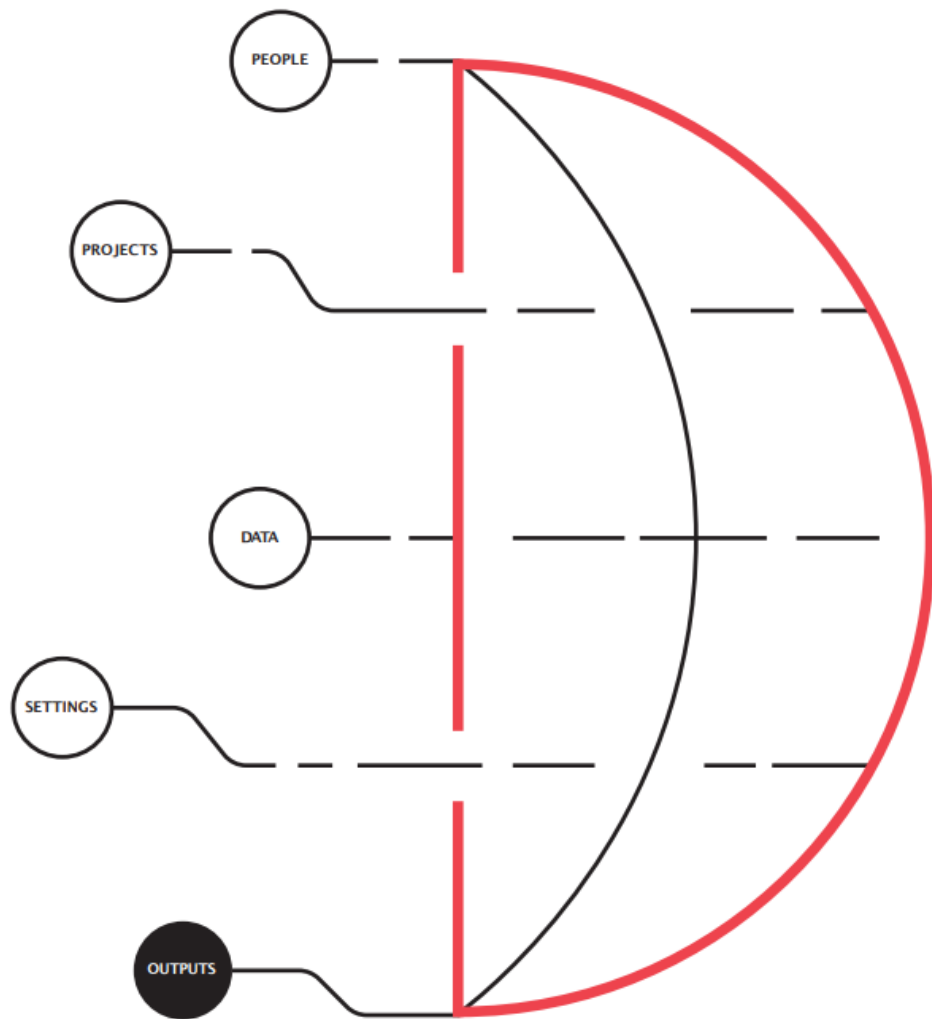
Not theory-heavy



Cross disciplinary



Free and publicly available



SECTION A

15

The statistical risk: what is it all about?

16

What is Statistical Disclosure Control?

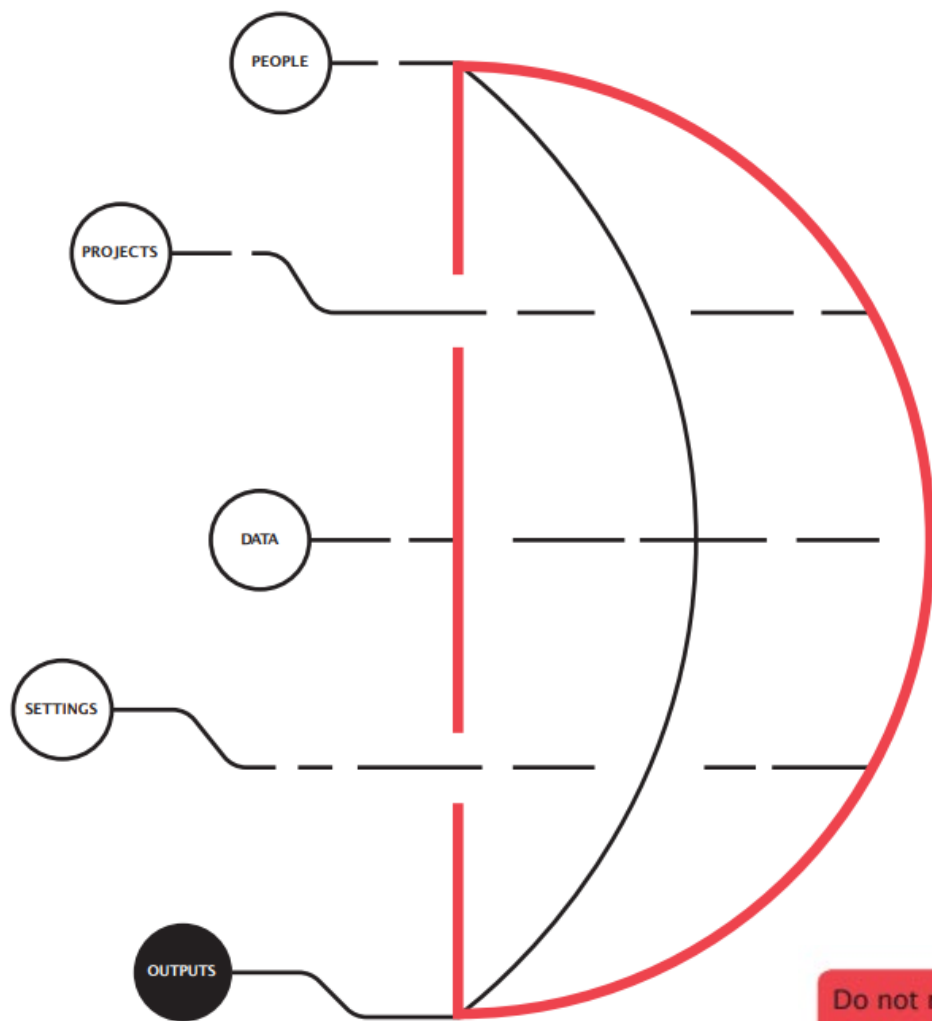
19

Risk assessment

22

The statistical risk: principles and rules

26



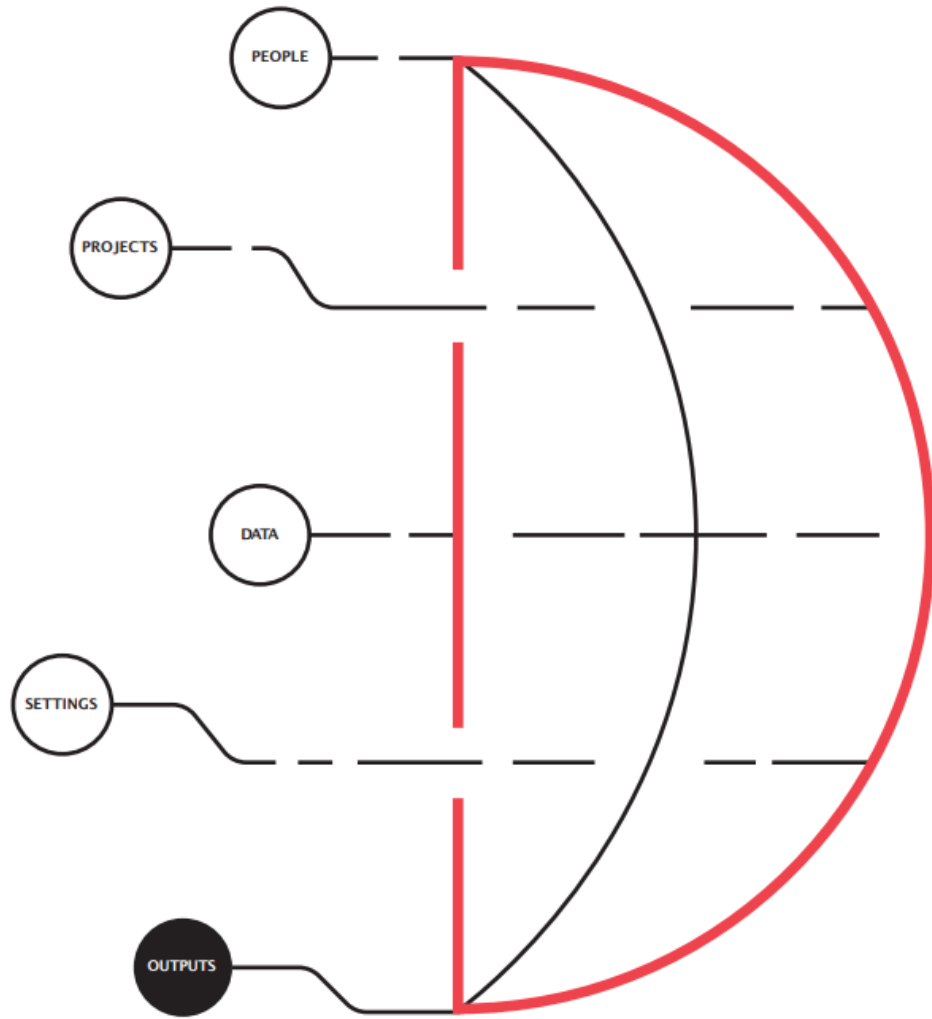
SECTION B

| | | | |
|--------------------------|----|---------------------------------------|----|
| Introduction: SDC | 28 | Symmetry plots | 48 |
| Using this guide | 29 | Decision trees and exclusion criteria | 51 |
| Descriptive statistics | 30 | Survival analysis: Kaplan–Meier curve | 53 |
| Percentiles | 31 | Spatial analysis (maps) | 55 |
| Histograms | 33 | Gini coefficients | 57 |
| Box plots | 35 | Concentration ratios | 59 |
| Correlation coefficients | 38 | Regressions | 61 |
| Factor Analysis | 40 | Residuals | 64 |
| Indices | 42 | Margin plots | 66 |
| Scatter plots | 44 | Test statistics | 68 |
| | 46 | | |

Do not release unless absolutely sure it's safe to do so

Might be safe to release, if a few sdc techniques are applied

Generally ok to release without much assessment



SECTION C

69

Introduction

70

Implementing SDC as an organisation

71

Managing analysts

76

What is a good set of statistical results?

78

The good

81

The bad

82

Managing expectations

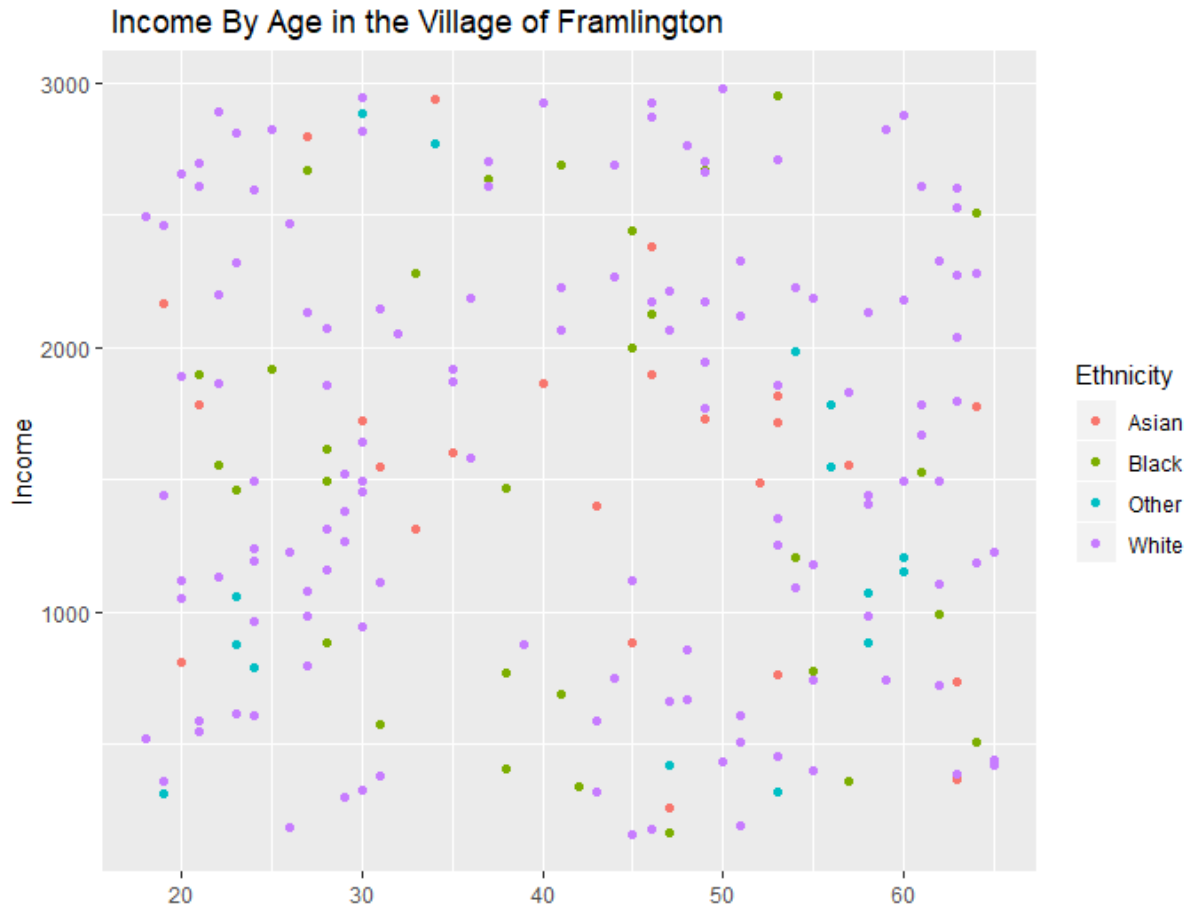
83

Audience participation

- Scatterplots
 - Does each point represent an individual person?
 - Do the variables plotted contain actual or perturbed information?
 - Would we reveal any confidential information by publishing the plot?
 - Could anybody be re-identified?

- On the next slide, we'll show you an example

Example



Does this scatter plot present a disclosure risk?

Publication and next steps

- Available to download at <https://securedatagroup.org/sdc-handbook/>
- FEEDBACK WELCOME!!
- Assessments and training – in trial right now
- Training materials based on the handbook
- E-learning exercise to follow in 2020

- Richard.welpton@health.org.uk
 - @rwelpton
- Simon.parker@health.org.uk
 - @SimonCRUK