

Spoken Statistical Data Summaries

Paul Piwek
Centre for Research in Computing



1

Aim: summarize statistical data automatically

Why? For people with visual impairment, inexperienced graph readers, users of mobile technology, ...

Development methodology: 1) Collect data and corresponding target texts. 2) Extract mappings from data to text and implement these in a system. The mappings can be created manually or (semi-)automatically using Machine Learning.

Data 1.1 (Source: ONS)

x-axis: year; **y-axis:** number of people aged 50 or over in the UK
data: (1901, 5678K), (1911, 6784K), (1931, 10442K), (1951, 13845K), (1961, 15960), (1971, 17300K), (1981, 17858K), (1991, 17974K), (2003, 20000K), (2011, 21957K), (2021, 25537K), (2031, 27213K)

Data 7.14 (Source: ONS)

x-axis: year; **y-axis:** pensioners' gross incomes in the UK in £
data: (1987, 193.42), (1988, 201.26), (1989, 195.93), (1990/91, 209.42), (1992, 218.48), (1993, 217.61), (1994/95, 225.12), (1995/96, 220.74), (1996/97, 234.57), (1997/98, 238.63), (1998/99, 246.44), (1999/00, 253.95), (2000/01, 269.10), (2001/02, 277.13), (2002/03, 280.41), (2003/04, 290.56)

Text 1.1 The data is about the number of people aged 50 and over in the UK. It gives the number by year, from 1901 until 2031. The figures for 2009 until 2031 are projections. In 1911, there were about 5 million people aged 50 or over. The number has steadily increased, with over 27 million projected for 2031.

Text 7.14 The data is about pensioners' income in the UK. It shows gross weekly income in pounds by year, from 1987 until 2003/2004. Gross income was £193.42 in 1987. The number increased to £290.56 in 2003/2004. There was a brief period, between 1988 and 1989, when gross income decreased from £201.26 to £195.93.

3

Meta-data

Goal

Input data

2

Raw data analysis

Salient patterns, chunks

Message selection and organization

Message templates

Selected and ordered messages

Text, prosody and gesture specification

Text with mark-up

