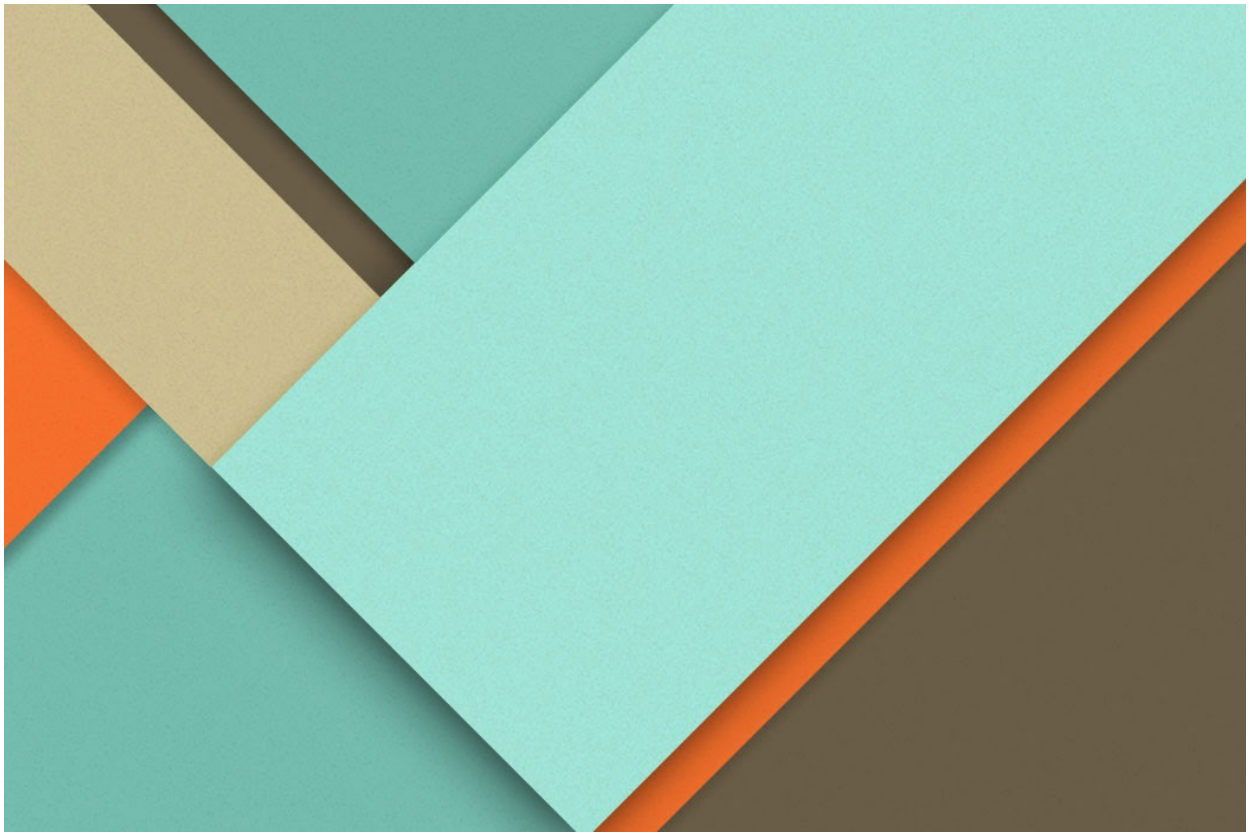

Health Project Description

12th August 2021

Group members:

- Barry O'Dea
- Paul Bateman
- Anne Braae
- Ben Irvine

Roles & responsibilities of each member

Barry worked on sourcing and cleaning the data, building graphs and the app design.

Paul worked on creating the file templates for the project description and the presentation.

Anne worked on sourcing data and creating a wireframe of the app, geospatial analysis and statistical research.

Ben worked on the initial Git repo setup, project oversight and graphs.

Everyone worked on planning; setting goals; assigning roles, prioritising tasks, version control of the app, and creating the presentation.

Brief description of dashboard topic

Our dashboard contains an overview of Scotland's health profile in the last 5-10 years, as well as specific insight into the effect on physical health of access to green spaces/parks. In addition, the app contains a visual illustration of the effect of neighbourhood deprivation on life expectancy.

Our health topic is the effect on physical health of access to green spaces/parks.

The dashboard outlines our health topic in terms of the breakdown by local authority area, and access to green space for each area. This is compared to the various physical health indicators, for the same areas (e.g. incidence of heart disease).

Stages of the project

1. Planning & dashboard wireframe
2. Data collection
3. Choosing datasets
4. Cleaning the data
5. Git branching & version control
6. Coding the R Shiny app framework
7. Exploratory data analysis

8. Creating basic visualisations and graph selection
9. Fine tuning the aesthetics elements of the app and content
10. Documenting and presenting our findings

Which tools were used in the project

- Zoom (daily stand-ups and occasional mob programming)
- Trello (planning & task allocation)
- Git/GitHub (collaboration & version control)
- Slack (team communication for written comments)
- Figma (wireframe of the R Shiny app)
- R (data-cleaning and wrangling)
- Canva (presentation)

Process considerations

How did you gather and synthesise requirements for the project?

We synthesised the information given in the brief by discussing the team's common interpretation of the brief, wire-framing our app design, and creating tasks to deliver the target outcome.

We prioritised the planning stage, data collection and build of a functional (MVP) application. This involved using Trello and wireframe discussions.

The team held many discussions on data suitability and selected the key datasets which enabled us to address the brief, keeping supplementary datasets for further analysis at a later stage.

Motivations for using the data you have chosen

We used the dataset on Scottish health survey to answer the requirement to give an overview of Scottish health in recent years because it contained various health indicator information by time, age group and region. We used data zone datasets to facilitate our spatial visualization of the health indicator data.

General question - status of exercise related health indices in Scotland?

We used the dataset on Distance to Green or Blue Space and the Scottish Health Survey Scotland Local Level Detail to ask the specific question:

Does convenient access to open space have an impact on measures of general well being health and exercise for the different local authority areas within Scotland.

List of datasets used:

Scottish health Survey- Scotland level data 2008-2019 - to get indicators of general well being, health and exercise in Scotland.

Scottish Health Survey Scotland. - Local Level Detail - to get indicators of general well being health and exercise in local areas of Scotland.

Distance to Green or Blue Space - Scottish Household Survey - to get indicators of distance from blue or green space.

Data Zone lookup - links geo-coordinates of the datazones to higher level areas.

Local Authority Boundaries - Scotland - spatial data frame for geographic statistics.gov.scot.

Life expectancy in years, at birth and for age groups.

- ## Data quality and potential bias, including a brief summary of data cleaning and transformations

Looking at the About tab for the main dataset on Scottish Health Survey data, the data quality control measures for the data included regular quality control checks at each stage of the survey. These included checks for the sampling design, selection of sampling methods, design of the questionnaire, how the survey was delivered, how the data was collected, entered and cleaned.

The health survey dataset may be biased because it contains self-reported data on various health metrics. Additionally, the survey did not capture older members of the population who do not live in private houses.

To clean the datasets we identified which data we needed to retain to complete our data analysis. For the health surveys we removed any health indicators that we felt were unnecessary to answer the questions we wanted to address. For the local area survey and greenspace data we removed the health board data so that we could concentrate on one geographical parameter and avoid duplicating analysis. We also merged in the location data so that we could have a local authority name associated with each area code. We finally renamed the headings in the datasets to make them uniform across the dataset we were comparing.

Data sources and ethical considerations

- ## How is the data on statistics.gov.scot stored and structured?

  Data on statistics.gov.scot is stored in the form of linked data. Each data point has a unique identifier URL which can be referred to by other linked data sets.

  These linked data points on statistics.gov.scot are stored as tri-part objects, triples, made of a subject, joined by a predicate to an object.

  Many predicates and objects can be joined to the same subject and can be used to complete the data known about a particular subject.

  What this looks like in RDF (Resource Description Framework, the "standard model for data interchange" online [source]) is a list of triples with different predicates and objects all with the same subject. The subjects within the data set can be joined to other subjects by a predicate, each subject being the "object" for each other.

  What this means is that the structure of the database is stored within the data and there is no database schema.

  Benefits of storing data like this are that it is very good at joining to other datasets and that it is easy to provide URLs for the data sources.

  [Further information]

- ## Ethical and legal considerations of the data

  There are no obvious ethical considerations, because the data does not identify particular individuals and their health details. There may be some misgivings among those citizens which have been categorised as living in a deprived area, perhaps for reasons of stigma and civic pride in their local area.

  The datasets are covered by the Open Government License, which means the data can be used for private analysis by members of the public, but should not be reproduced for commercial gain.