Aim

The aim of this report was to highlight the difference in mental wellbeing scores for male and female high school students in S2 and S4 in the council areas of Scotland in 2012. The reduction in the mean mental wellbeing scores from S2 to S4 was larger for girls than for boys. It was also noted that the lowest scores for S4 girls were in more rural areas of Scotland. The intended audience is local government, schools, and parents. These groups could provide support and help to girls in S4 and focus resources in more rural areas of Scotland.

Why is this issue important?

The most recent Executive Summary of the Scottish Schools Adolescent Lifestyle and Substance Use Survey (SALSUS), 2021, stated that low mental health wellbeing scores were found to be higher in pupils living in high deprivation areas, with disabilities and caring responsibilities [1]. It was also highlighted that 15 year old girls (S4 high school) have had the lowest wellbeing scores since 2010. Governments, schools and parents should be doing everything they can to support all pupils with low mental wellbeing scores. The pupils' education could be seriously affected, resulting in a lack of job offers in the future. Also providing less skilled people to improve the economy of Scotland [2].

How can data science help?

There were multiple websites, versions of reports and data files available on the SALSUS research. The vast amount of data could be overwhelming for the public. Moreover, the important key health message may be unclear. It is important for data scientists to make important findings, such as the example above, easily accessible to the public. This can be done by researching the topic in detail, sourcing the correct most relevant information, questioning any anomalies, and highlighting the key points.

The research undertaken can be used as a catalyst to create new government policy and practice [3]. The findings from this report could result in guidelines being created for schools to better support S4 girls. Also, the resources available to S4 Girls in all council areas could be utilised to really help those in need [4]. Funds could be targeted towards the areas that have the lowest mean wellbeing scores.

This initial key finding could also result in further research. Why do some council areas have lower mean wellbeing scores than others? Perhaps population density could be used to compare the mean wellbeing score to determine if there is a relationship. Any new findings would benefit the affected population and their families. It may even lead to new tools being designed, such as an App, that S4 girls could use to monitor their wellbeing [5]. If Big Data was collected using this new tool in addition with users' medical records, data scientists could utilise the data to best determine how this group could be supported [6].

What data was used?

The mean mental wellbeing scores for boys and girls in S2 and S4 in Scotland, 2012, were sourced from the Online Profile Tools available from Scottish Public Health Observatory [7]. This data had a reference to the SALSUS report.

The first limitation is that the data is now out of date. Also, information about the 2012 survey could not be found, however the survey from 2013 stated that only 60% responded [8]. If this was also the case for the 2012 the survey could have non-response bias because all the pupils didn't complete the survey. The students that didn't complete the form could be off school due to mental health issues or illness. These students would be likely to have low mean wellbeing scores. This would result in a higher mean wellbeing score, from the survey, as many low scores will not have been accounted for. Information about which council areas the non-response surveys were from was not available. There might be a particular council that did not provide data. Further investigation would be required.

The 2013 SALSUS report also stated teachers issued the survey. This could result in performance bias unless strict guidelines and instructions were followed by each teacher. If some students were given less time to complete the survey or unclear instructions this could result in inaccurate data.

The mean wellbeing scores were calculated using the Warwick-Edinburgh Mental Wellbeing Scale (WEMWBS) [9]. The answers from each survey will have to be processed in the same manner. If the data from the surveys is recorded incorrectly then systematic bias could occur. There is no information about how the data has been cleaned or checked for errors. Not all analysts would use the same method to clean the data, resulting in different outcomes [10]. The data was available on the website as secondary data, caution must be taken when using this type of data.

The map shapefiles for the data zone boundaries, dated 2011, used in creating the map were sourced from the Scottish Government [11]. The data zone information for each council area was sourced from the Scottish Index of Multiple Deprivation 2020v2 data [12]. The joining of these datasets along with the mean wellbeing score data for S4 girls (2012) were used to create the map. The years of the data joined were 2011, 2020 and 2012 however the data zones and council areas should be the same as the year 2012, so this is not a concern.

Strengths and Limitations

Column Chart

A column chart was chosen instead of a pie chart even though there were only 4 observations to display. This is because the human eye finds it difficult to detect changes in angles [13]. The differences in the mean wellbeing score would not have been noticeably different using angles.

The colours of the columns make it clear that it is for male (blue) and female (pink) participants. However, this may be offensive to some people because it can be seen as stereotypical. If the research had provided non-binary data this could have been added in a third column, using a neutral colour. The colours chosen might not be clear to people who are colour blind. Therefore, even though the legend was removed the labels of each bar retained the information of the gender.

The measure variable (mean wellbeing score) could have been added as a label to the top of each column. Reading an accurate value from the y-axis due to the continuous numeric scale would be difficult. However, it is only the general trend that is needed for the intended audience. The actual

mean wellbeing score may not be meaningful to the intended audience. An understanding of how the mean wellbeing score was calculated and how to interpret the score would be required.

An improvement to the column chart would be to create a group 2 column chart [14]. This would have resulted in a gap between the bars of the two genders. This would have highlighted the difference more clearly. It was also important to highlight the decreasing mean wellbeing scores, so therefore rather than comparing the high school years by gender the gender was compared by high school year. This highlighting the decline in mean wellbeing score for girls towards the right of the graph.

The difference in mean wellbeing scores could have been emphasised by changing the start of the y-axis value to a non-zero value. This however could result in miss leading the intended audience. The lower mental wellbeing score for S4 girls can be seen clearly in the graph, but it is not clear on how much this difference is affecting the lives of pupils between the different groups.

Map

Unfortunately, not all the council areas in Scotland participated in the 2012 SALSUS survey. Therefore, there was missing data for some of the council areas. The pale not applicable areas could be misleading because this lighter colour may be interpretated as having a high wellbeing score, unless the reader carefully refers to the legend.

When initially loaded the interactive map has zoomed out and only by zooming in closer are the different council areas clear. It is possible to click on the different council areas and the name and mean wellbeing scores are then visible.

Different palettes of colours were used to represent the mean wellbeing scores. A palette was chosen that clearly showed a darker colour, to represent a darker mood, and brighter colour to represent a happier score. The map colours are unique for each of the chosen variable values. In this case it is a continuous scale. Initially code was written to define a colour for each of the 24 council areas, but this resulted in a very large legend. Therefore break points were used which only required 7 colours. The legend was then much easier to understand.

The actual value of the mean wellbeing score for each council area can be found by clicking on the interactive map. Interactive resources have been found to make data more open and accessible [15] Initially the popup attributes had information that was not relevant, such as the type of data (multipolygons) and the data zone information. This would not be of interest to the intended audience. The popup was changed to only include the council area and mean wellbeing score. The feature ID is still visible, but with further work it is possible to change this too.

A summary of the information could have been added as a table. This would have required packages such as flexdashboard and gt. However, there are multiple council areas so the table would potentially distract the reader from the map. Each council has multiple data zones, so how the scores vary within the towns and cities was not available within the data set. Therefore, large geographical areas have only one mean wellbeing score. However, overall the map has potential to convey a clear message about the mental wellbeing of S4 girls in different council areas in Scotland.

References

1. Scottish Schools Adolescent Lifestyle and Substance Use Survey (SALSUS): mental wellbeing report 2018. [cited 2022 Dec 5]; Available from: https://www.gov.scot/publications/scottish-schools-adolescent-lifestyle-substance-use-survey-salsus-mental-wellbeing-report-2018/

2. 5 health and work: The Scottish picture. [cited 2022 Dec 5]; Available from: https://www.gov.scot/publications/fair-healthy-work-review-scottish-governments/pages/5/

3. Subrahmanya SVG, Shetty DK, Patil V, Hameed BMZ, Paul R, Smriti K, et al. The role of data science in healthcare advancements: applications, benefits, and future prospects. Ir J Med Sci [Internet]. 2022;191(4):1473–83. Available from: http://dx.doi.org/10.1007/s11845-021-02730-z

4. Why is health data important? — [Internet]. Data Saves Lives. [cited 2022 Nov 25]. Available from: https://datasaveslives.eu/why-is-health-data-important

5. Public Health Scotland. Data modelling tool supports quality delivery of mental health services [Internet]. Publichealthscotland.scot. [cited 2022 Dec 5]. Available from: https://publichealthscotland.scot/our-blog/2022/october/data-modelling-tool-supports-quality-delivery-of-mental-health-services/

6. Pastorino R, De Vito C, Migliara G, Glocker K, Binenbaum I, Ricciardi W, et al. Benefits and challenges of Big Data in healthcare: an overview of the European initiatives. Eur J Public Health [Internet]. 2019;29(Supplement_3):23–7. Available from: http://dx.doi.org/10.1093/eurpub/ckz168

7. Online Profiles Tool [Internet]. Org.uk. [cited 2022 Nov 26]. Available from: https://www.scotpho.org.uk/comparative-health/profiles/online-profiles-tool/

8. SALSUS National Overview 2013 [Internet]. Available from: https://www.isdscotland.org/Health-Topics/Public-Health/Publications/2014-11-25/SALSUS_2013_National_Overview.pdf

9. Marmara J, Zarate D, Vassallo J, Patten R, Stavropoulos V. Warwick Edinburgh Mental Well-Being Scale (WEMWBS): measurement invariance across genders and item response theory examination. BMC Psychol [Internet]. 2022 [cited 2022 Dec 6];10(1):31. Available from: https://bmcpsychology.biomedcentral.com/articles/10.1186/s40359-022-00720-z

10. Silberzahn R, Uhlmann EL, Martin DP, Anselmi P, Aust F, Awtrey E, et al. Many analysts, one data set: Making transparent how variations in analytic choices affect results. Adv Methods Pract Psychol Sci [Internet]. 2018;1(3):337–56. Available from: https://psyarxiv.com/qkwst

11. Scottish Government SpatialData [Internet]. gov. scot. Data Zone Boundaries 2011. 2021. [cited 2022 Nov 26]. Available from: https://www.data.gov.uk/dataset/ab9f1f20-3b7f-4efa-9bd2-239acf63b540/data-zone-boundaries-2011

12. Scottish Index of Multiple Deprivation 2020v2 - indicators. [cited 2022 Nov 26]; Available from: https://www.gov.scot/publications/scottish-index-of-multiple-deprivation-2020v2-indicator-data/

13. PROTECT benefit-risk [Internet]. Protectbenefitrisk.eu. [cited 2022 Nov 25]. Available from: https://protectbenefitrisk.eu/piecharts.html

14. Zach. How to create a grouped barplot in R (with examples) [Internet]. Statology. 2020 [cited 2022 Dec 6]. Available from: https://www.statology.org/grouped-barplot-in-r/

15. Perkel JM. Data visualization tools drive interactivity and reproducibility in online publishing. Nature [Internet]. 2018 [cited 2022 Dec 3];554(7690):133–4. Available from: https://www.nature.com/articles/d41586-018-01322-9