Data Explained

Suspected Cancer Pathway Monthly (SCPM)

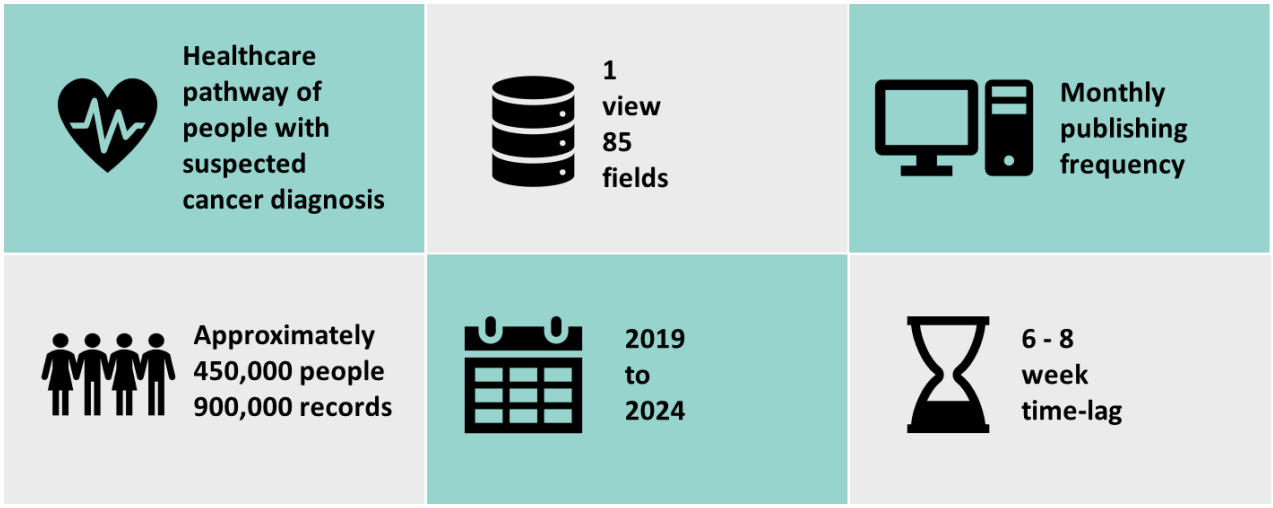
Authors: Dr Laura Baker, Dr Arron Lacey, Professor Ashley Akbari
Date: February 2025

This Data Explained summarises the Suspected Cancer Pathway (Monthly (SCPM)) data. This output is intended to help guide users interested in this data for the first time, and future use of the data within the SAIL Databank towards research delivery and to provide feedback into future data source development and documentation.

The data discussed in this Data Explained was made securely available through an approved SAIL project via the independent Information Governance Review Panel (IGRP) – project 1598. The data used in this approved SAIL project comes from Digital Health and Care Wales (DHCW) and was accessed through the Secure Anonymised Information Linkage (SAIL) Databank. The data was not originally collected for research and it is expected that there are gaps and inconsistencies in its recording, a number of which are detailed in the following. The work presented in this Data Explained is correct at the point of publication. Views expressed in this Data Explained are those of the researchers and not necessarily those of ADR Wales partner organisations.

Overview

The Suspected Cancer Pathway Monthly (SCPM) data source is accessible through the SAIL Databank and relates to individuals undergoing diagnostic testing when suspected of having cancer and covers all recorded cancer services in Wales. The infographic below summarises the data sources' features at the time of this publication.



Introduction

Cancer is a complex and significant public health issue, being a leading cause of death globally (Siegel et al., 2023; WHO, 2020). The increasing incidence of cancer, combined with an ageing population, places greater demand on healthcare services and highlights the need for effective detection, treatment, and intervention strategies.

The SCPM data source relates to individuals undergoing diagnostic testing when suspected of having cancer and covers all recorded cancer services in Wales. The Welsh Government implemented a Suspected Cancer Pathway (SCP) in Wales in June 2019, with the purpose of setting a target for more rapid cancer diagnosis and subsequent treatment provision, in addition to identifying areas within the pathway that require additional support (*Suspected Cancer Pathway, 2021*). Similarly, this data is captured in England via the National Cancer Waiting Times Monitoring Data Set, (found here), in Scotland via the Cancer Waiting Times in Scotland Data Set (found here) and Cancer Wait Times Northern Ireland (found here).

From its design, the SCPM was initially split into urgent and non-urgent cancer pathways, and these were designated as National Statistics (Welsh Government, 2023). As such the SCP based cancer statistics were described as experimental, until a review in 2022 concluded that the methods were appropriate and the statistics were robust and reliable (Welsh Government, 2023).

The SCPM contains all individuals that have undergone an investigative treatment, which may or may not lead to a cancer diagnosis (*Suspected Cancer Pathway, 2021*). This includes individuals

identified with cancer and also those without. The data contains individual level cancer pathway information such as pathway timestamps, demographic data and data validity checks.

How is the data collected?

When an individual is suspected of having a cancer diagnosis (clock start, as defined by *NHS Data Dictionary SCPM Scope*, 2024), the National Health Service (NHS) report an expectation that the individual under investigation is tracked from the point of suspicion until the treatment date or a cancer diagnosis is ruled out (*NHS Data Dictionary SCPM Scope*, 2024).

As such, data is collected via the SCPM by healthcare providers for any individual suspected of having a new primary cancer, regardless of previous diagnoses (including Squamous Cell Skin Cancer, which is excluded in registry data (Public Health Wales NHS Trust, 2024). Data is then submitted for all pathways (open or closed). The NHS SCPM scope states: "*All patients who receive their first definitive treatment within Wales should be included…. Patients who are referred by secondary care for treatment outside of NHS Wales will be included in the reporting…., those that enter directly or are referred from primary care directly to services outside of Wales will not.*" (*NHS Data Dictionary SCPM Scope*, 2024).

Health boards collect data daily on individual cancer services via the Welsh Patient Administration System (WPAS) or local systems. Extracts of SCP are submitted from local health boards to DHCW monthly, which is then validated and aggregated. Following this the aggregated data is then submitted to the Welsh Government between the fourth and fifth working day of each month. There is approximately a six-week lag from the reference month (Welsh Government, 2023). For comprehensive details on data validation, compilation, and adjustment processes, the Welsh Government has produced an in-depth guide that outlines each of these procedures in detail (Welsh Government, 2023).

Data Linkage

Accessing SCPM through the SAIL Databank allows cancer diagnostic and outcomes data to be linked with other data sources, including routinely-collected electronic health record (EHR), administrative, survey, registry and other data sources. A full list of data sources available via the Databank is provided here: https://healthdatagateway.org/en/search?type=datasets&sampleAvailability=&publisherName=SAIL.

This data linkage supports the development of more specific and targeted research questions. For example, it enables studies on how time between suspected cancer and diagnosis date may affect staging of disease, or analysis on the impact of diagnostic timeframes vary across geographical locations, in addition to many other ways which linkage could be used to evaluate trends at the population level.

What can the data be used for?

SCPM is a single view within the SAIL Databank, that contains 85 fields (additional information surrounding the data structure can be found at HDR UK, 2024).

Linkage is facilitated to other data sources via the Encrypted Anonymised Linking Field (ALF) field, which serves as a double encrypted unique identifier linked to individuals in Wales.

Currently there are 873,167 records, for 424,547 individuals (**Table 1**), covering a timeframe of 2019 -2024 within the SCPM data source. The increase in the number of records compared to the number of individuals may be due to various factors. For instance, a single individual might have multiple suspected primary cancers, or there may be missing data at the ALF (Anonymous Linkage Field) level.

Table 1. The number of records, individuals and missing individuals detailed within the SCP data source. Counts were taken in 2024, and represent the entire data source.

Description	Count
Total Records	873,167
Total Individual Records	834,706
Number of Missing ALF	38,461
Number of Individuals	424,547

The fields within the SCPM data provide insight into the following:

1. **An individual's diagnostic journey**: The data facilitates tracking an individual's pathway from initial referral or suspicion of cancer through to diagnosis and treatment. In SCPM, 27 out of 85 fields are dates or timestamps, relating to events such as diagnostic tests, referrals, and where applicable, the date of death.

2. **Data validation fields:** There are 33 fields designated for data validation in SCPM, offering binary true/false checks to ensure data accuracy. These fields indicate whether the data has been cross-referenced with additional records, with a true value signifying that the data has been validated.

3. **Demographics:** Demographic information is provided, including week of birth, sex, and ethnic code, alongside health care interaction codes and geography, such as local health board, and details of tertiary treatment centres.

Potential Uses of the Data

The SCPM data can serve as a resource for incorporating additional parameters into population-based cancer studies focused on disease detection and diagnosis. For example, it enables the calculation of metrics such as the average time to key events. Specifically, this includes:

- Average wait time from the first diagnostic test to diagnosis (11 days).
- Average time from pathway start to cancer diagnosis (33 days).
- Average time from pathway start to tertiary referral (53 days).
- Average time from pathway start to date of death (90 days).

These metrics are calculated based on individual patient episodes, with the time to each event averaged across the data and rounded to the nearest day. Please note that in-depth analysis of outliers has not been conducted, as it was beyond the scope of this Data Explained.

Using SCPM as a standalone data source, it is possible to analyse the geographical distribution of individuals entering the SCP across local health boards over time (**Figure 1**). This analysis enables targeted interventions and resource allocation if higher concentrations of individuals are observed in specific areas.

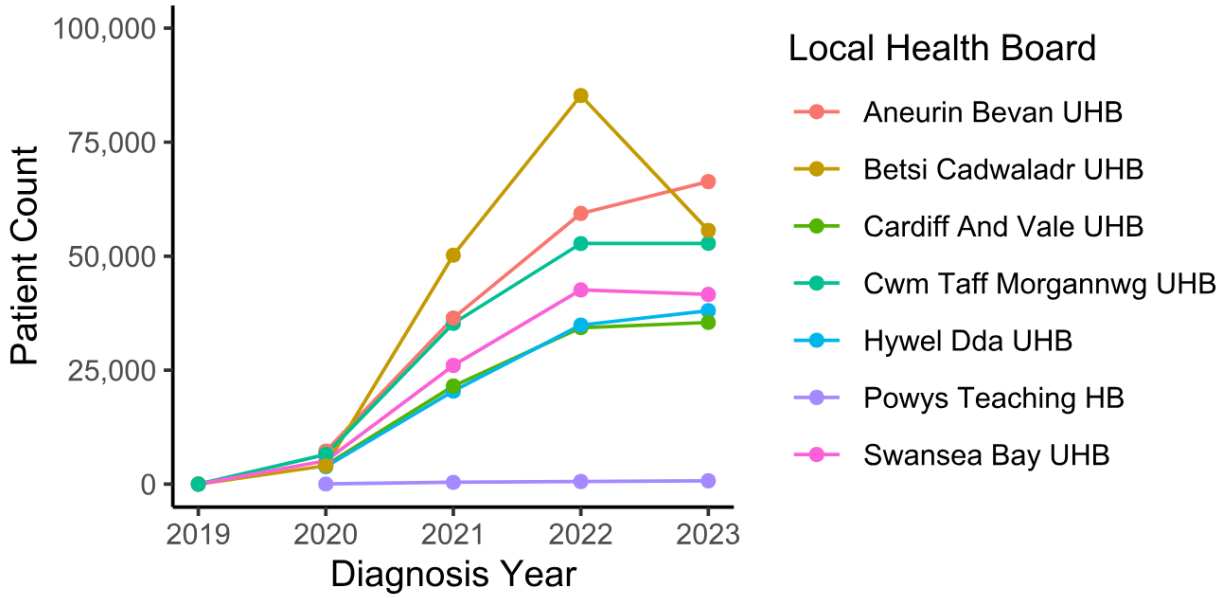


Figure 1. A longitudinal overview of individuals entering the SCP per local/university health board (UHB/HB) in Wales.
Individual's counts are based on unique entries within the SCPM data source, representing distinct patient episodes. * *Please note that the patient count for Powys Teaching Health Board is displayed in this visual, but it appears as zero due to the comparatively larger counts for other health boards. For a more detailed breakdown of each health board, please refer to Appendix Figure 1.*

SCPM data also supports the analysis of how many individuals enter the SCP compared to those who receive a primary cancer diagnosis each year (**Figure 2**). Given that SCPM is a relatively new data source and healthcare services worldwide have been affected by events like the COVID-19

pandemic, it is difficult to assess the utility of yearly comparisons between the two groups. However, this analysis provides valuable insights into the current data coverage, highlighting a growing number of entries over time (2019 – 2023).

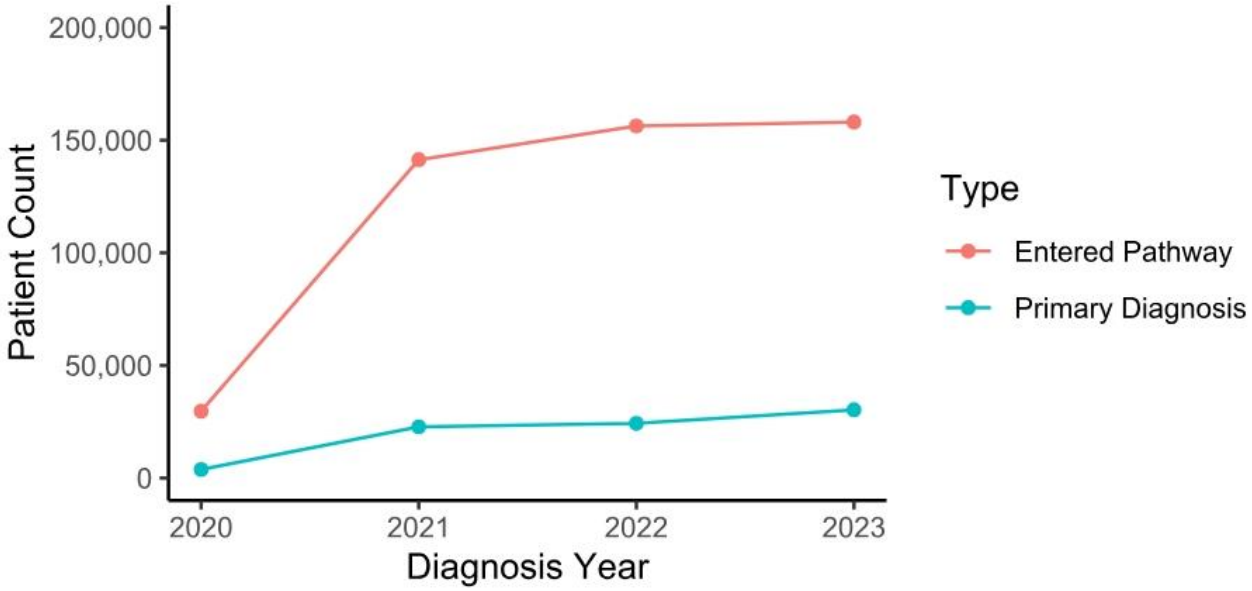


Figure 2. A longitudinal comparison of suspected cancer (red line), and cancer diagnosis (blue line) in Wales. Counts were conducted on distinct patient episodes and grouped per year; cancer diagnosis date/pathway start date.

In addition to the traditional annual counts, future analysis could track changes over time in the number of individuals diagnosed with cancer versus those who are not. Other potential insights might include the average age of individuals entering the SCPM and various other relevant metrics.

Existing research or examples of previous research

SCPM was introduced in Wales in 2019 and is therefore a relatively new data source. Consequently, there are currently no studies specifically reporting on the use of the SCPM data. However, several studies have utilised similar data from hospitals or other data repositories to investigate cancer wait times and inform different practises (Tang et al., (2023). For example, Scott et al, (2023) examined future cancer risk based on urgent suspected cancer referral data, Round et al., (2020) studied the urgency of referrals and related outcomes, and Abelardo et al., (2022) investigated the incidence of referrals.

Data limitations encountered and suggested improvements

Overall, the documentation for SCPM is well described within the data dictionary, and the provision of a user guide from the Welsh Government provides context for data collection and validation (Welsh Government, 2023). However, there are several factors that should be considered when evaluating the data:

- **Data validation:** While the validation fields support data integrity, the reliance on cross-referencing external records may present gaps if these records are incomplete or unavailable. In addition, there is no reference relating to the process of this validation system.
- **Missing lookup tables:** The absence of lookup tables for certain fields, hinders interpretation and analysis, necessitating external resources or assumptions to derive insights.
- **Newness of data source:** Given that the data is relatively new, the short coverage period makes long-term trend analysis challenging.

In addition, a notable limitation of relying solely on SCPM data is that it only captures individuals suspected of having cancer. This excludes cases where individuals do not enter the SCP but are later diagnosed with cancer under different circumstances, such as posthumous diagnoses.
This data holds significant potential for analysing individual cancer pathways but may require additional context or external references to fully realise its value.

Summary

Overall, SCPM provides valuable insights into an individual's cancer journey when combined with cancer registration data. It enables retrospective trend analysis of the timescales experienced by individuals who had, or were suspected of having, cancer. For example, it may be possible to identify trends in individuals with longer wait times, stage of cancer at diagnosis and ultimately prognosis. Understanding these factors can have significant implications for healthcare services, highlighting areas needing attention, potential improvements, or aspects that are performing well. Utilising SCPM can offer crucial insights to drive policy action and enhance cancer care in Wales.

References

Abelardo, E., Gravelle, R., Scannell, M., Shastri, P., Vandekar, M., Davies, G., Volpini, L., & Prabhu, V. (2022). Impact of coronavirus disease 2019 on head and neck urgent suspected cancer referral pathways in rural Wales. *The Journal of Laryngology & Otology*, *136*(6), 540–546. https://doi.org/10.1017/S002221512200069X

HDR UK. (2024). *Suspected Cancer Pathway Monthly (SCPM)*. https://web.www.healthdatagateway.org/dataset/fcf5a844-f48d-428e-aa06-abe64e66185e

Lyons, R. A., Jones, K. H., John, G., Brooks, C. J., Verplancke, J.-P., Ford, D. V., Brown, G., & Leake, K. (2009). The SAIL databank: Linking multiple health and social care datasets. *BMC Medical Informatics and Decision Making*, *9*(1), 3. https://doi.org/10.1186/1472-6947-9-3

NHS Data Dictionary SCPM Scope. (2024). NHS Wales Data Dictionary. https://www.datadictionary.wales.nhs.uk/index.html#!WordDocuments/scope83.htm

Public Health Wales NHS Trust. (2024). *WCISU Cancer Reporting Tool Technical Guide*. https://phw.nhs.wales/services-and-teams/welsh-cancer-intelligence-and-surveillance-unit-wcisu/cancer-reporting-tool-official-statistics/

Round, T., Gildea, C., Ashworth, M., & Møller, H. (2020). Association between use of urgent suspected cancer referral and mortality and stage at diagnosis: A 5-year national cohort study. *British Journal of General Practice*, *70*(695), e389–e398. https://doi.org/10.3399/bjgp20X709433

Scott, S. E., Gildea, C., Nicholson, B. D., Evans, R. E., Waller, J., Smith, D., Purushotham, A., & Round, T. (2023). Future cancer risk after urgent suspected cancer referral in England when cancer is not found: A national cohort study. *The Lancet Oncology*, *24*(11), 1242–1251. https://doi.org/10.1016/S1470-2045(23)00435-7

Siegel, R. L., Miller, K. D., Wagle, N. S., & Jemal, A. (2023). Cancer statistics, 2023. *CA: A Cancer Journal for Clinicians*, *73*(1), 17–48. https://doi.org/10.3322/caac.21763

Suspected Cancer Pathway. (2021, January 5). NHS Wales Executive. https://executive.nhs.wales/functions/networks-and-planning/cancer/workstreams/suspected-cancer-pathway/

Tang, A., Chandler, S., Torkington, J., Harris, D., & Dhruva Rao, P. (2023). Adapting the investigation of patients on urgent suspected cancer pathway with lower gastrointestinal symptoms across Wales during COVID-19. *The Annals of The Royal College of Surgeons of England*, *105*(S2), S35–S41. https://doi.org/10.1308/rcsann.2021.0366

Welsh Government. (2023, August 10). *Suspected cancer pathway: Quality report | GOV.WALES*. https://www.gov.wales/suspected-cancer-pathway-quality-report-html

WHO. (2020). *The top 10 causes of death*. World Health Organisation. https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death

Glossary

Abbreviation	Name
ADR	Administrative Data Research UK
ALF	Anonymised Linkage Field
DHCW	Digital Health and Care Wales
ESRC	Economic and Social Research Council
IGPR	Information Governance Review Panel
NHS	National Health Service
PHW	Public Health Wales
SAIL	Secure Anonymised Information Linkage (Databank)
SCPM	Suspected Cancer Pathway/ Monthly
TRE	Trusted Research Environment
WCRC	Wales Cancer Research Centre
WCSU/WCISU	Welsh Cancer Intelligence and Surveillance Unit
WPAS	Welsh Patient Administration System

Disclaimer

This work was produced using administrative data accessed through The SAIL Databank. The use of the data in this work does not imply the endorsement of SAIL or data owners in relation to the interpretation or analysis.
This work uses research data source which may not exactly reproduce National Statistics aggregates. National Statistics follow consistent statistical conventions over time and cannot be compared to Data First linked data source.

Acknowledgements

This work is supported by ADR Wales, part of ADR UK (Administrative Data Research UK), a partnership transforming access to public sector data in the UK to inform policy decisions that enhance people's lives. ADR UK is funded by the Economic and Social Research Council (ESRC), part of UK Research and Innovation ES/W012227/1]. The research was facilitated by the SAIL Databank, with funding from Health and Care Research Wales via the Wales Cancer Research Centre's CReSt catalytic funding (researcher salary and data access costs) and funding from the Roche Bioresource Data Accelerator programme (data access costs).

We gratefully acknowledge the contributions and support of Professor Sunil Dolwani, Professor Mererid Evans, Dr Mark Davies, Liz Merrifield, Dr Peter Giles and Jenni Macdougall.

Contact

Name: Laura Baker
Email: l.m.baker@swansea.ac.uk

Appendix

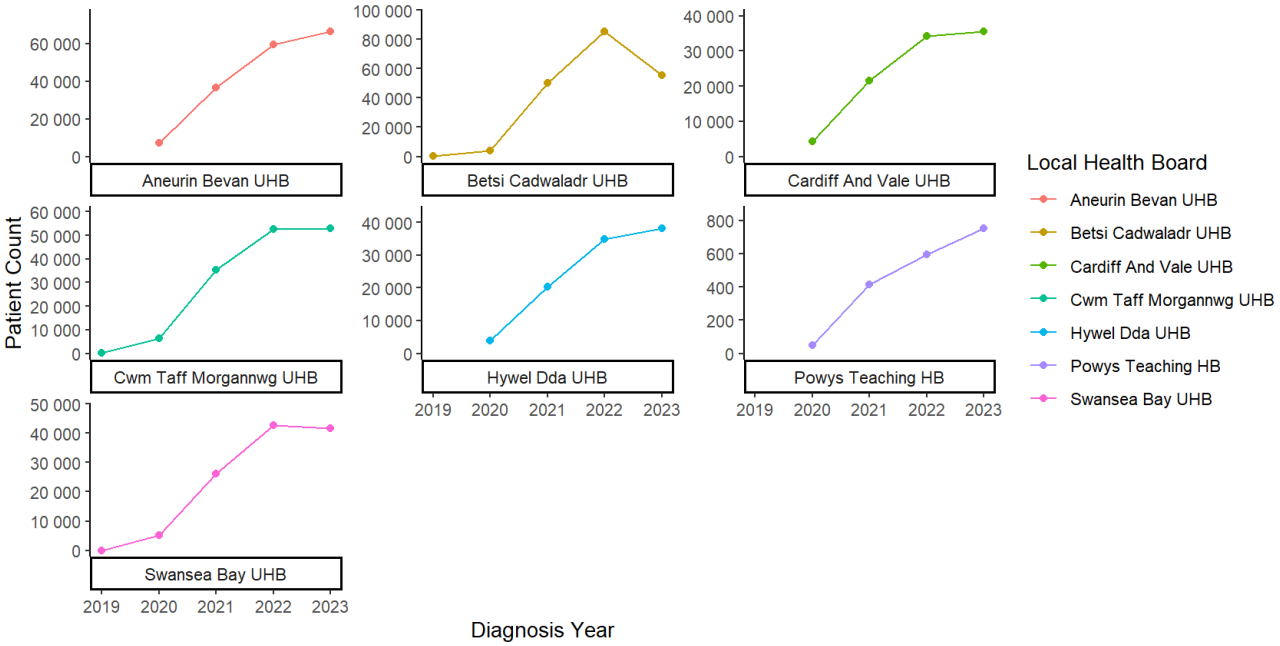


Figure A1. A longitudinal overview of individuals entering the suspected cancer pathway per local/university health board (UHB/HB) in Wales. The visual is presented as a faceted plot, allowing for more detailed counts to be clearly observed.