



Islas Canarias
Del 15 al 19 de noviembre de 2021



Panel de datos de Turismo de Lanzarote: Una aplicación de datos abiertos en la administración pública.

Christian González-Martel

Métodos Cuantitativos en Economía y Gestión, Universidad de Las Palmas de Gran Canaria. Las Palmas, España.
christian.gonzalez@ulpgc.es

José Manuel Cazorla-Artiles

Economía de las Infraestructuras, el Transporte y el Turismo (EITT)
Departamento de Análisis Económico Aplicado, Universidad de Las Palmas de Gran Canaria, Las Palmas, España.
josemanuel.cazorla@ulpgc.es

Introducción

Durante los últimos años, las administraciones públicas han dado pasos significativos para ofrecer en abierto sus datos. Thorsby et al. (2017) investigaron el contenido y características de portales de datos abiertos de ciudades de Estados Unidos. Los resultados mostraron que, en general, los portales estaban en fases iniciales de desarrollo y necesitaban mejorar la ayuda de usuario y las herramientas de análisis, así como incluir apartados que ayudasen a los usuarios entender los datos, más gráficos y análisis. Ruijter et al. (2017) concluyeron que diseños basados en datos abiertos ayudaban a la transformación de los datos en crudo a información útil.

Por otro lado, en el ámbito académico existe una tendencia a que la investigación publicada sea reproducible, esto es, que los investigadores permitan la consulta de los códigos y los datos usados en el estudio. Esto implica que el código debe estar disponible, los datos accesibles en formato legible y debe existir una plataforma que permita su distribución. (Stodden et al., 2014) Cada vez más revistas de investigación obligan a los autores a compartir, al menos los datos, en sus publicaciones. Los flujos de trabajo usados en la investigación que abarcan desde los datos en bruto hasta la publicación final, pueden beneficiarse de la disponibilidad de herramientas algorítmicas que permitan acceder y analizar datos abiertos (Boettiger et al., 2015; Gandrud, 2013).

Sin embargo, debido a que los datos proporcionados en acceso abierto no están en un formato estandarizado, éstos deben adaptarse para adaptarse a las necesidades del usuario. El objetivo es que un programador, que debe pasar los datos en crudo a información usable, evite tareas de conversión lo que permite la simplificación, estandarización y automatización de los flujos de trabajo. Esto redundaría en los objetivos antes señalados de reproducibilidad, intercambio de código y análisis de datos eficiente. Siguiendo esta idea,

dentro del ecosistema de R, se han creado varios paquetes para trabajar con datos proporcionados por la Organización de las Naciones Unidas para la Agricultura y la Alimentación (FAO) [FAOSTAT (Kao et al., 2015)], Banco Mundial [WDI (Arel-Bundock, 2021), wbstats (Piburn, 2020)], Open Street Map [osmar (Eugster & Schlesinger, 2012)] entre otros.

Objetivos

El ISTAC es el órgano central del sistema estadístico autónomo y centro oficial de investigación del Gobierno de Canarias. Entre sus funciones se encuentran proporcionar información estadística y coordinar la actividad estadística pública de la región autónoma de Canarias. Proporciona una amplia colección de datos, que incluye conjuntos de datos sobre la demografía, la salud, el empleo y el turismo de Canarias y otros temas en un formato de datos abiertos.

El ISTAC ha hecho un esfuerzo para que sus datos sean accesibles. La principal es a través de una interfaz web donde se pueden consultar y descargar los datos en distintos formatos. En ocasiones, este método de acceso es tedioso para grandes selecciones y cuando el usuario debe acceder a los datos con mucha frecuencia. Otro método alternativo utiliza una interfaz de programación de aplicaciones (API) que se puede incrustar en un código para extraer los datos de forma programática.

El objetivo de este trabajo fue desarrollar un paquete R que integrase la API del ISTAC y permite manipular directamente los datos descargados en el software estadístico R.

Como segundo objetivo, se planteó el desarrollo de una aplicación Shiny que usase los datos obtenidos con este paquete y que permitiese visualizar y analizar los datos del ISTAC relacionados con los indicadores del turismo en Lanzarote. En esta aplicación, la forma de presentar la información es principalmente a través de gráficos. La visualización es una de las características más importantes en el análisis de información de fuentes de datos abiertas. Jones et al. (2016) investigaron mecanismos innovadores de visualización e intercambio de datos en el estudio de datos de encuestas de ciencias sociales sobre cuestiones ambientales para permitir la deliberación participativa.

Este trabajo proporciona una descripción general del paquete en su versión actual. Una documentación más completa, así como código fuente están disponibles en Github (<https://github.com/rOpenSpain/istacbaser>). El paquete forma parte de rOpenSpain, una iniciativa cuyo objetivo es crear paquetes de R para explotar los datos abiertos disponibles en España para una investigación reproducible.

Este artículo está estructurado de la siguiente manera: en primer lugar, explicamos el procedimiento de obtención de datos implementado en la librería de R y la rutina para lograr la visualización de los datos descargados. En el apartado de Resultados explicamos la arquitectura del Tablero de Indicadores de Turismo de Lanzarote.

Finalmente, presentamos las conclusiones del trabajo.

Metodología

Descarga y visualización de datos con istacbaser

El paquete desarrollado utiliza la API del ISTAC ‘ISTAC BASE’ y lleva por nombre ‘istacbaser’. Para instalar y cargar la última versión de la librería, el usuario debe escribir en R el comando de instalación desde GitHub desde el paquete ‘remotes’.

```
remotes::install_github("rOpenSpain/istacbaser")
library("istacbaser")
```

Cuando se carga el paquete, los metadatos de cada conjunto de datos disponible mediante ISTAC BASE API se cargan en la variable ‘cache’. Contiene información sobre el título, tema, subtema, la url para acceder a los datos json, entre otros. Para buscar un término específico, se proporciona la función de búsqueda `istacbase_search`.

```
busqueda.paro <- istacbase_search("paro",
                                  fields = "datos publicadosI")}
```

Este comando busca entre todos los conjuntos de datos proporcionados por la API ISTAC BASE aquellos en los que el patrón de búsqueda ‘paro’ aparece dentro del campo “datos publicados I”. Otros campos pueden ser “título” (predeterminado), “tema”, “subtemaI”, “subtemaII”, “datos publicados I”, “origen” y “encuesta”. Puede obtenerse la lista de campos con `names(cache)`.

El patrón de búsqueda se puede utilizar con operadores de expresiones regulares. La salida de la función `istacbase_search` son las filas o fila de `cache` que cumplen con los requisitos del patrón de búsqueda. Los valores de la columna ID de la salida proporcionan identificadores que se usarán para la posterior descarga de los datos. Por ejemplo, podemos filtrar el ID del conjunto de datos que recopila el desempleo registrado según sexo y grupos de edad por islas y meses usando.

```
selectedID <- busqueda.paro %>%
select(titulo, ID) %>%
filter(grepl("sexo", titulo) &
       grepl("edad", titulo) &
       grepl("Islas", titulo)) %>%
pull(ID)

## [1] "emp.est.na.ser.4866"
```

Descarga de datos de ISTAC

Recuperamos los datos del conjunto de datos con la referencia de ID utilizando la API ISTAC BASE.

```
df <- istacbase(selectedID)
```

Por defecto, la función `istacbase` trabaja con etiquetas legibles. Con el argumento `label = FALSE`, la función transforma las etiquetas en códigos menos legibles.

Los indicadores con los que trabaja el ISTAC suelen estar disponibles como series de tiempo anuales agrupadas por islas, pero a veces con una granularidad o niveles geográficos diferentes. Si el conjunto de datos tiene la columna “islas”, se puede filtrar por islas usando el argumento `islas = TRUE`; de lo contrario, este argumento se ignora. Los valores válidos para ‘islas’ son: El Hierro, La Palma, La Gomera, Tenerife, Gran Canaria, Fuerteventura y Lanzarote.

La función permite filtrar el conjunto de datos por fechas utilizando los argumentos `startdate`, `enddate` y `mrv`. El argumento `freq` controla la granularidad de los datos. Se

puede obtener valores anuales ("anual"), semestrales ("semestral"), trimestrales ("trimestral"), mensuales ("mensual"), quincenales ("quincenal"), semanales ("semanal").

Visualización de datos

'Istacbaser' por sí solo no tiene una función dedicada para representar datos, pero puede usar el potencial que proporciona R para visualizar los datos obtenidos. La Figura 1 muestra el resultado de la combinación del paquete istacbaser y ggplot2

```
ggplot (df %>% filter(`Grupos de edad` == "TOTAL" &
                    Sexos != "AMBOS SEXOS"),
aes(x = fecha, y = valor, colour = Sexos)) +
geom_line() +
facet_wrap(~Islas, scales="free_y") +
theme_light()+
theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
labs(x = "Years",
     y = "Registered unemployment",
     title = "Registered unemployment according to sex and islands,
     subtitle = "May 2005 - July 2021")
```

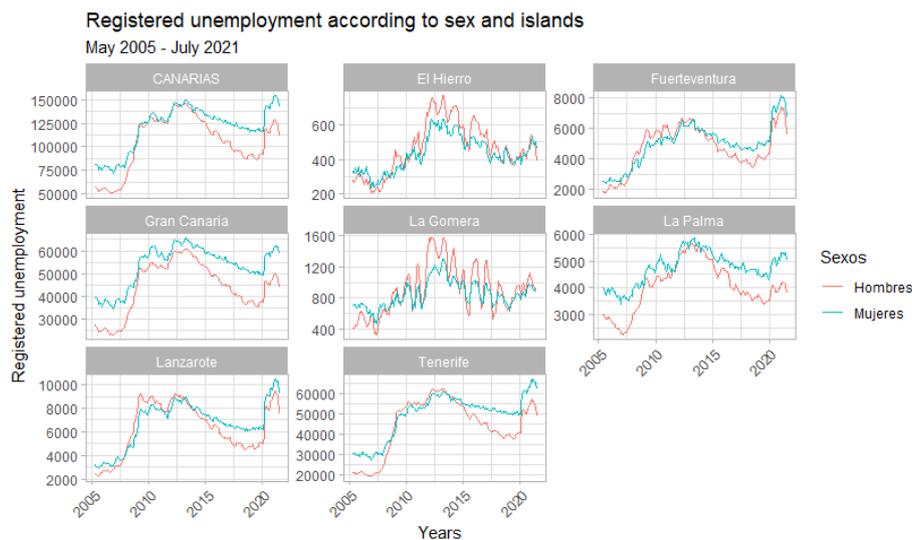


Figura 1: Datos obtenidos a través de istacbaser. Visualización con función ggplot del paquete ggplot2. Fuente: ISTAC (2018)

Resultados

Panel de Datos de Turismo de Lanzarote

Con el fin de proporcionar información relacionada con el turismo de la isla Lanzarote, el Centro de Datos de Lanzarote, entidad asociada al Cabildo de Lanzarote que facilita información estadística de Lanzarote, y la Universidad de Las Palmas de Gran Canaria establecieron una colaboración para desarrollar un portal web basado en Shiny que mostrase los indicadores de turismo relevantes de manera amigable a los usuarios de la plataforma.

Se puede acceder a la aplicación a través de <https://datosdelanzarote.shinyapps.io/turismo-dashboard/>.

El panel se basa en la idea del Panel de Turismo de Nueva Zelanda. En ese sentido, se desarrolló el Panel de Indicadores de Turismo de Lanzarote. El código está escrito en software estadístico R con un uso preponderante del paquete Shiny [16], que se utiliza como framework de R para uso de javascript. El aplicativo funciona bajo servidores en shinyapps. El Panel de Indicadores de Turismo de Lanzarote permite al usuario descargar datos y gráficos en formato png. Las principales visualizaciones de esta aplicación web se realizan con los paquete dygraphs , billboarder y leaflet.

El paquete istacbaser es fundamental en el desarrollo de esta aplicación porque es el conector entre los datos de ISTAC y el Panel de Indicadores de Turismo de Lanzarote. Todos los datos se gestionan con este paquete. La librería istacbaser obtiene la información que previamente ISTAC ha obtenido mediante encuestas. El Tablero de Indicadores de Turismo de Lanzarote explota estadísticas que provienen de tres encuestas realizadas por el ISTAC: Encuesta de Gasto Turístico (EGT), Encuesta de Movimientos Turísticos en Fronteras (FRONTUR) y Encuesta de Ocupación en Alojamientos Turísticos.

El fin de la aplicación Shiny es mantener informados a los actores relacionados con el turismo de Lanzarote sobre la situación del sector. La información proporcionada por el Panel de Indicadores de Turismo de Lanzarote abarca tanto la oferta y la demanda turística en Lanzarote.

Por parte de la demanda se analiza las llegadas de turistas, gastos realizados en destino, características del viaje y perfil del turista.

La evolución de las llegadas desde los principales mercados de origen se muestra mediante diferentes gráficos de series temporales y la información proviene principalmente de la Encuesta de Movimientos Turísticos en Fronteras (FRONTUR). La evolución se puede observar según una periodicidad mensual o anual. El usuario también puede ver los valores en nivel, como promedio mensual de 12 meses o como variación anual. Existen varias opciones para analizar los resultados (Figura 2), según el número de noches, el mercado de origen o el tipo de alojamiento.

Llegadas de turistas

Llegadas de turistas a Lanzarote según países de residencia

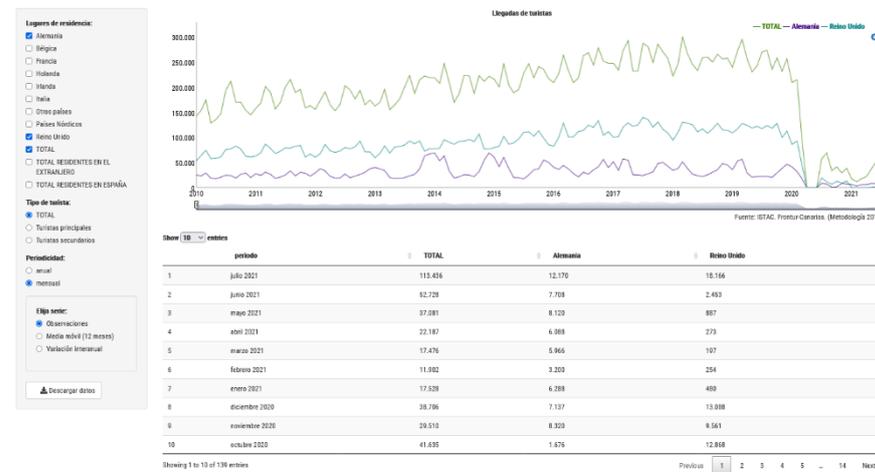


Figura 2: Llegadas de turistas. Panel de Indicadores de Turismo de Lanzarote. Acceso: 2021-10-09

La información del gasto se centra en el gasto total, el gasto por turista y el gasto por turista y día. La evolución del gasto de los principales mercados de origen proviene de la Encuesta de Gasto Turístico. Puede analizarse con frecuencia anual o trimestral. Existen varios filtros para analizar el mercado de origen incluso con un desglose por partidas de gasto.

La información proporcionada por las características del viaje está enfocada en la valoración de aspectos para elegir el destino, el tipo de alojamiento, la duración de la estadía y las actividades que realizan los turistas. La evolución de las características de los viajes de los principales mercados de origen proviene de la Encuesta de Gasto Turístico. La evolución podría analizarse con frecuencia anual o trimestral

El perfil del turista también se analiza para describir las características del turista en términos de género, edad, ocupación e ingresos. Esta información proviene de la Encuesta de Gasto Turístico. La evolución de los turistas según estas características podría analizarse con gráficos de series temporales con frecuencia anual o trimestral. Asimismo, se desarrolló un mapa de viaje del cliente para describir por período y mercado de origen las características y el comportamiento del turista antes y durante el viaje y la valoración después del viaje (Figura 3).

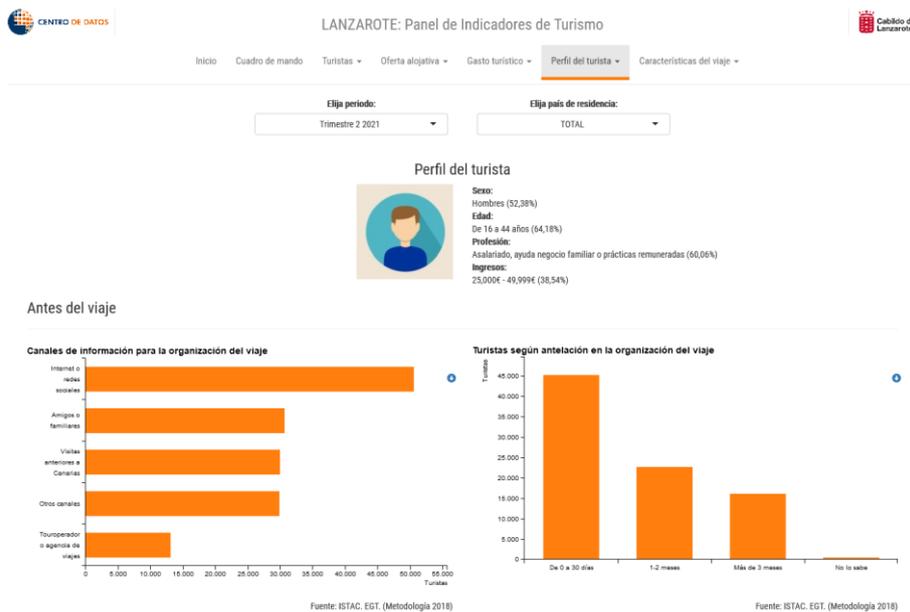


Figura 3: Perfil del turista. Panel de Indicadores de Turismo de Lanzarote. Acceso: 2021-10-09.

Por el lado de la oferta, se utiliza como indicadores la ocupación, la capacidad de alojamiento, los ingresos y el empleo. La información de oferta proviene de la Encuesta de Ocupación de Alojamientos Turísticos. La evolución de la oferta turística se puede analizar con frecuencia anual o mensual, con observaciones en nivel o variación anual para las diferentes regiones de Lanzarote, según el tipo de alojamiento o el mercado de origen. El lado de la oferta muestra indicadores, tales como, número de establecimientos, número de habitaciones, tasas de ocupación, tarifa diaria promedio (ADR), ingresos totales y empleo turístico entre otros (Figura 4).

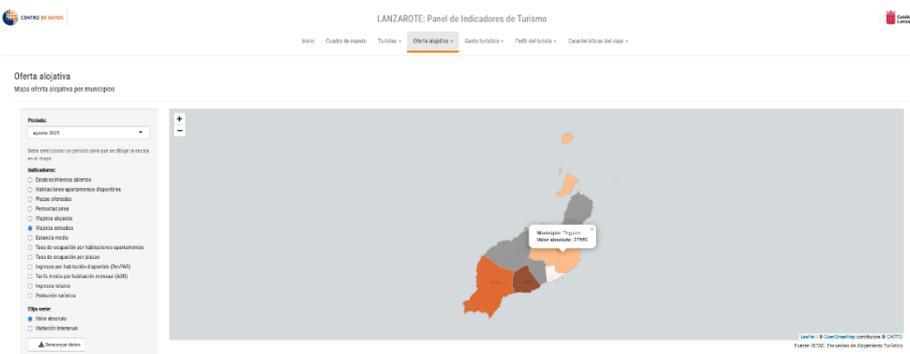


Figura 4: Oferta alojativa. Panel de Indicadores de Turismo de Lanzarote. Acceso: 2021-10-09

Dentro de la sección “Cuadro de mando” (Figura 5) el usuario puede ver de un vistazo la situación del sector turístico en Lanzarote por periodo. Los indicadores de oferta y demanda turística se resumen en una sola pestaña. En concreto, los indicadores mostrados son llegadas, gasto por turista y día, duración de la estancia, tasa de ocupación, número de noches y ADR. Asimismo, existen gráficos para mostrar el desglose de llegadas por

mercado de origen y gráficos de series temporales para la evolución de llegadas, ADR y gasto por turista y día.

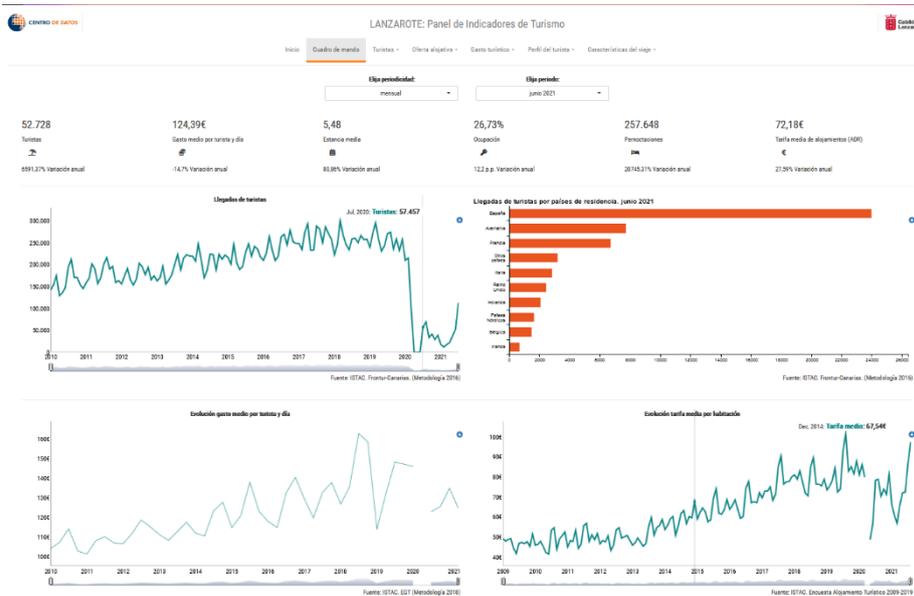


Figura 5: Cuadro de mando. Panel de Indicadores de Turismo de Lanzarote. Acceso: 2021-10-09.

Conclusiones

Cada vez son más las administraciones públicas que están tomando conciencia de la necesidad de que sus datos estén en abiertos como compromiso para con los ciudadanos. Esto es útil para investigadores y expertos para el desarrollo de proyectos que usen estos datos pero la heterogeneidad de los formatos, su complejidad y actualizaciones continuas obligan al uso de herramientas computacionales que traten estos datos así como herramientas de visualización que muestre la información contenida en estos datos.

Los objetivos de este proyecto fueron principalmente dos. Primero, desarrollar una librería de R, *istacbaser*, que administrase conjuntos de datos accesibles por API ISTAC BASE. Este paquete permite manejar los datos proporcionados por ISTAC programáticamente para uso directo o para incluirlos en otros desarrollos.

En segundo lugar, crear una aplicación web para visualización de los datos proporcionados por ISTAC BASE. El Cabildo de Lanzarote a través del Centro de Datos de Lanzarote ha financiado el proyecto donde se aplica este desarrollo. El resultado es el Panel de Indicadores de Turismo de Lanzarote.

En este sentido, la aplicación web es una mejora frente a los informes estáticos. El Panel permite al usuario interactuar con la aplicación y obtener diferentes respuestas según la configuración establecida. Esto es bastante útil en el caso de usuarios no especializados porque no necesitan conocimientos sobre programación o gestión de datos, y pueden obtener diferentes respuestas manejando solamente la configuración de la aplicación. Este servicio proporciona a los profesionales del sector toda la información relacionada con el turismo reunida en un solo lugar para su consulta y análisis, permitiendo así que la institución pública, en este caso el Centro de Datos de Lanzarote, logre uno de sus principales objetivos como es la “difusión de información estadística y documental sobre Lanzarote”.

Referencias

- Arel-Bundock, V. (2021). *World Development Indicators and Other World Bank Data [R package WDI version 2.7.4]*. <https://cran.r-project.org/package=WDI>
- Boettiger, C., Chamberlain, S., Hart, E., & Ram, K. (2015). Building Software, Building Community: Lessons from the rOpenSci Project. *Journal of Open Research Software*, 3(1). <https://doi.org/10.5334/JORS.BU>
- Eugster, M. J. A., & Schlesinger, T. (2012). *osmar: OpenStreetMap and R*. <http://api.openstreetmap.org/api/>
- Gandrud, C. (2013). Reproducible Research with R and R Studio. In *Reproducible Research with R and R Studio*. <https://doi.org/10.1201/b15100>
- Jones, A. S., Horsburgh, J. S., Jackson-Smith, D., Ramírez, M., Flint, C. G., & Caraballo, J. (2016). A web-based, interactive visualization tool for social environmental survey data. *Environmental Modelling & Software*, 84, 412–426. <https://doi.org/10.1016/J.ENVSOF.2016.07.013>
- Kao, M. C. J., Gesmann, M., & Filippo, G. (2015). *Package 'FAOSTAT'*. 22. <https://cran.r-project.org/web/packages/FAOSTAT/index.html>
- Piburn, J. (2020). *Programmatic Access to Data and Statistics from the World Bank API [R package wbstats version 1.0.4]*. <https://cran.r-project.org/package=wbstats>
- Ruijter, E., Grimmelikhuijsen, S., & Meijer, A. (2017). Open data for democracy: Developing a theoretical framework for open data use. *Government Information Quarterly*, 34(1), 45–52. <https://doi.org/10.1016/J.GIQ.2017.01.001>
- Stodden, V., Leisch, F., & Peng, R. D. (2014). Implementing reproducible research. In *Implementing Reproducible Research*. <https://doi.org/10.1201/b16868>
- Thorsby, J., Stowers, G. N. L., Wolslegel, K., & Tumbuan, E. (2017). Understanding the content and features of open data portals in American cities. *Government Information Quarterly*, 34(1), 53–61. <https://doi.org/10.1016/J.GIQ.2016.07.001>