Automatización de reportes (y un poco más) con R







Pablo Aguirre Hörmann

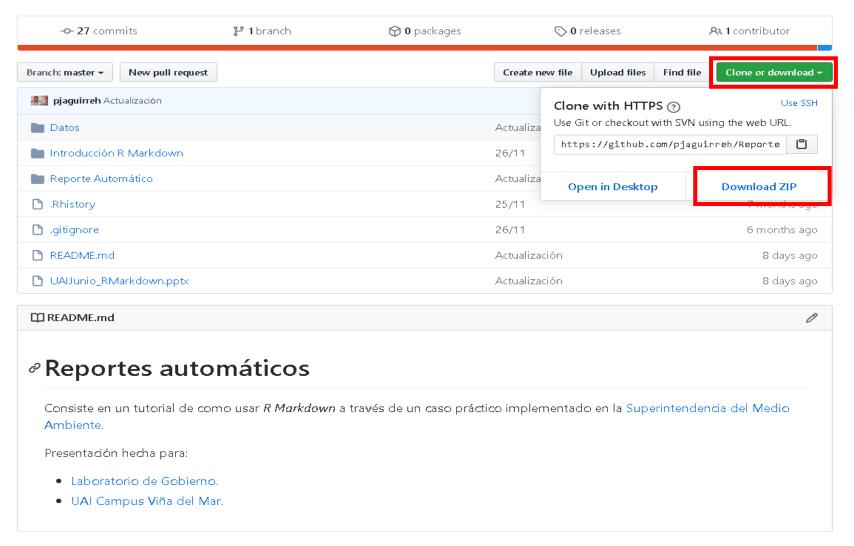
pjaguirreh@gmail.com

¿Qué veremos?

- 1. ¿Por qué automatizar?
- 2. ¿Qué es R, RStudio, y R Markdown?
- 3. Demostración de generación automática de reporte
- 4. Otras posibilidades con R (y otros software)

Este no es un taller para aprender a programar en R

Si tienes instalado R/RStudio GitHub



https://github.com/pjaguirreh/Reportes-automaticos

¿Por qué automatizar?

Ahorrar tiempo

- ¿Siempre?
- Tareas repetitivas

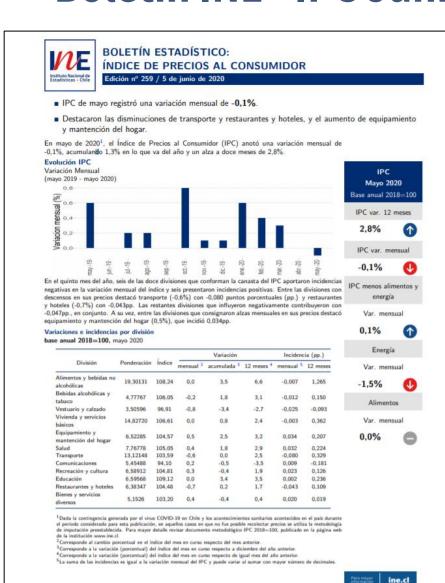
Minimizar errores

Y aparición de otros

Mejorar gestión

- Información actualizada y fácilmente disponible
- Posibilidad de hacer cosas que antes no se podían

Boletín INE - IPC Junio



Divisiones

División 7 de transporte



Transporte anotó disminuciones mensuales en tres de sus diez clases. La más importante fue combustibles y lubricantes para vehículos de transporte personal (-3.1%) que incidió -0.096pp, mientras que transporte de pasaieros por vías urbanas y carreteras (-0,3%), contribuyó con -0,007pp. De los 24 productos que componen la división, siete presentaron bajas en sus precios, siendo el más relevante gasolina (-3,0%), con una incidencia de -0,082pp., seguido de petróleo diésel (-4,1%), con -0,015pp. Los restantes productos con incidencias negativas acumularon -0,009pp.

División 11 de restaurantes y hoteles



Restaurantes y hoteles consignó descensos mensuales en una de sus dos clases (a un dígito). La más importante fue servicios de alojamiento (-8,8%) que aportó -0,042pp., seguido de restaurantes, cafés y establecimientos similares (0,0%), con -0,001pp. De los siete productos que componen la división, uno consignó bajas en sus precios (a un dígito), destacando servicio de alojamiento turístico (-8,8%), con una incidencia de -0,042pp., seguido de alimentos consumidos fuera del hogar (0,0%), con -0,001pp. Los restantes productos con contribuciones negativas acumularon -0,001pp.

División 5 de equipamiento y mantención del hogar



Equipamiento y mantención del hogar registró aumentos mensuales en nueve de sus diez clases. La más importante fue muebles y artículos para el hogar (1,0%) que aportó 0,010pp., seguido de grandes artefactos eléctricos o no para el hogar (0,9%), con 0,006pp. De los 36 productos que componen la división, 25 consignaron alzas en sus precios, destacando lavadora (3,7%), con una incidencia de 0,008pp., y muebles para living (2,3%), con 0,007pp. Los restantes productos con contribuciones positivas acumularon 0.033pp.

Productos

Gasolina

En mayo, presentó una disminución mensual de 3,0%, aportando -0,082pp., a la variación del indicador general. Acu- acumulando -0,3% en lo que va del año y -1,4% a doce meses. muló -5,4% al quinto mes del año y -2,1% a doce meses.

Carne de nollo

Anotó un descenso mensual de 5.2%, con una incidencia de -0,046pp., y variaciones de -4,7% en lo que va del año y de -1.4% a doce meses.

Servicio de alojamiento turístico

Registró una baja mensual de 8,8%, con -0,042pp., acumulando -11,3% en lo que va del año y -12,0% a doce meses.

Consignó una variación mensual de -1,7%, con -0,024pp.,

Carne de vacuno

Consignó un aumento mensual de 2,0%, con una incidencia de 0,039pp., registrando una variación de 6,8% en lo que va del año y de 12,0% a doce meses.

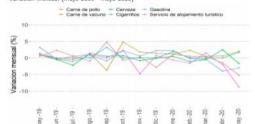
Cigarrillos

Registró un alza mensual de 1,7%, con 0,034pp., acumulando 3,6% en lo que va del año y 7,0% a doce meses.

Evolución productos

2

Variación mensual (mayo 2019 - mayo 2020)



ELIPC es utilizado como:

- Medida official de la inflacción en el cais.
- Determinante de la evolución de la UF y UTM Medida para ajustar o traer a valor presente contratos y/o montos.

Nuestra calculadora te permite ajustar monto Dempo en períodos que tú escajas (www.ine.cl)

Instituto Nacional de Estadística randé 801, piso 22, Santiago, Chil Teléfono (56) 23246 1132-1137



Indice de Precios al Consumidor Nº 259 - Boletín Estadístico del Instituto Nacional de Estadísticas

INE publica el primer boletín de IPC automatizado y reduce drásticamente tiempos de elaboración

El proceso permite, además minimizar en forma sustancial los riesgos de errores. Este es el segundo boletín que la institución automatiza, después del IPP, y se espera continuar con los boletines de empleo y de sectores económicos.

Como parte del objetivo de modernizar sus procedimientos y gracias a un trabajo de programación que tomó cerca de dos meses, el Instituto Nacional de Estadísticas (INE) publicó este mes, por primera vez, una versión automatizada de su boletín del Índice de Precios al Consumidor (IPC). De esta forma, el IPC pasó a ser el segundo producto estadístico en ser elaborado mediante este proceso, después de que el boletín del Índice de Precios de Productor (IPP) fuera automatizado en 2019.

Este procedimiento permite reducir sustancialmente el tiempo que toma elaborar esos boletines puesto que se trabaja sobre un diseño y un texto preestablecido y ya no son personas, sino que un programa informático, el que elabora un PDF del boletín directamente a partir de la base de datos. Ello ha significado, además, minimizar los riesgos de errores de digitación y de generación de gráficos o tablas. Los programas informáticos correspondientes fueron habilitados para esta tarea de automatización por integrantes del propio equipo del área de Estadísticas de Precios del INE.

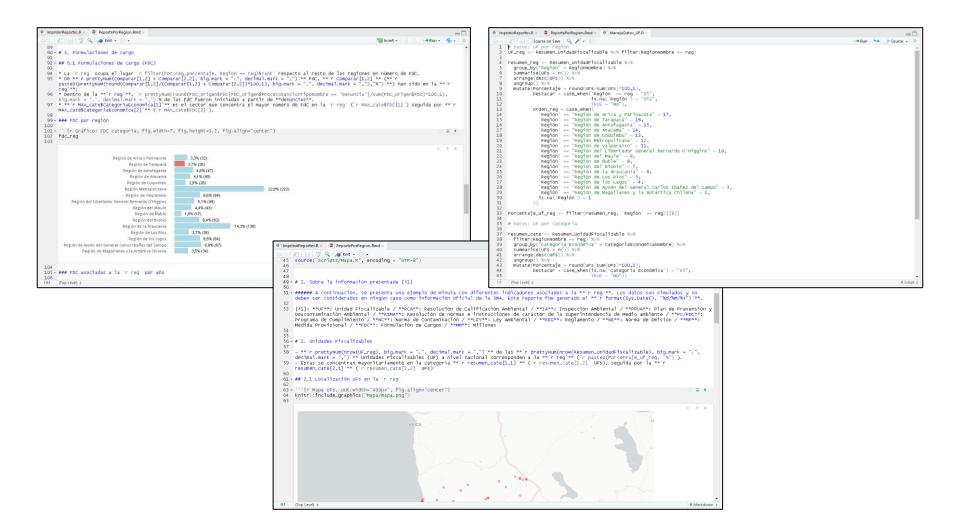
Si antes la elaboración final de un boletín como el IPC o IPP podía tomar hasta seis horas (incluyendo la redacción, la diagramación y las revisiones), ahora con la automatización el proceso requiere solo segundos y luego algunos minutos para corregir eventuales detalles de formato.

Minutas automatizadas

A lo que queremos llegar



Minutas automatizadas Por lo que tenemos que pasar





Lenguaje y plataforma

- Lenguaje y ambiente de programación estadística
- Herramienta de visualización de datos
- Gratuito

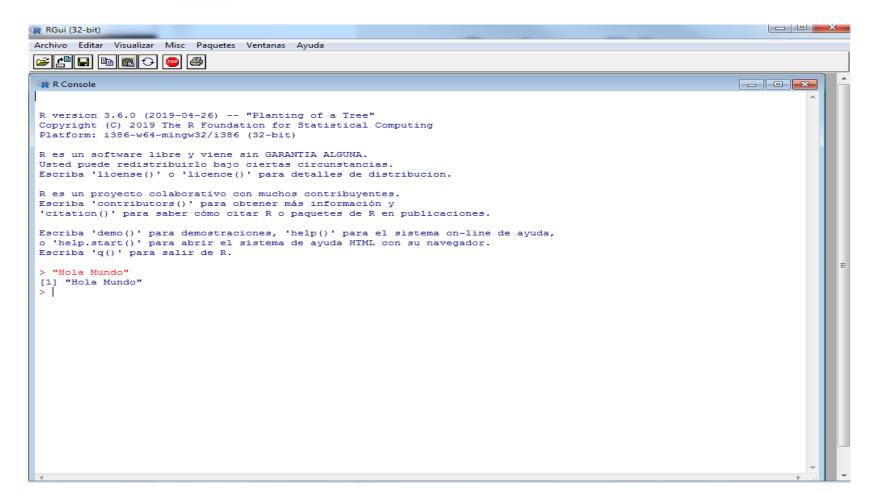
Ecosistema

- Muchas aplicaciones e integraciones con otras plataformas
- 15.000+ librerías gratuitas disponibles

Comunidad

- Millones de usuarios
- Muchos y diversos grupos de usuarios a nivel mundial







IDE para R: Entorno de desarrollo integrado

Consola

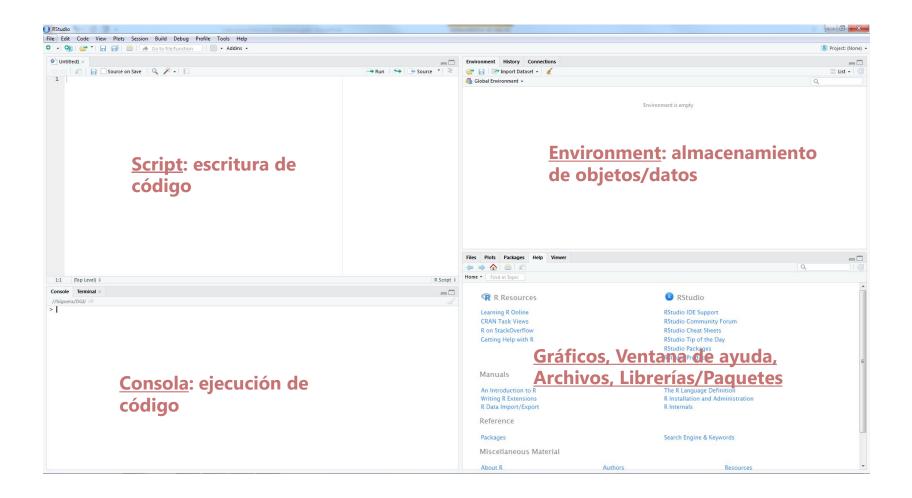
Editor de código

- Auto-completación
- Ayuda de sintaxis
- Ejecución directa

Herramientas para distintas tareas

- Visualización
- Conexión con otras plataformas
- Depuración de código
- Manejo del ambiente de trabajo

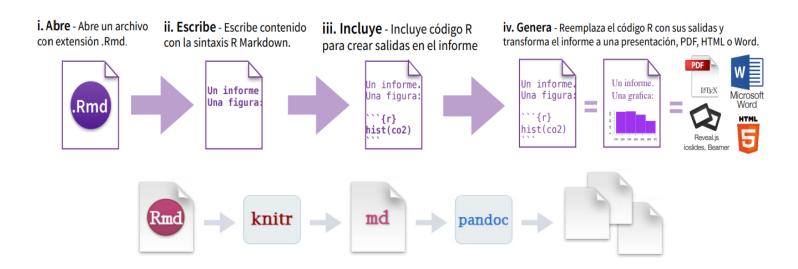
¿Qué es R Studio?



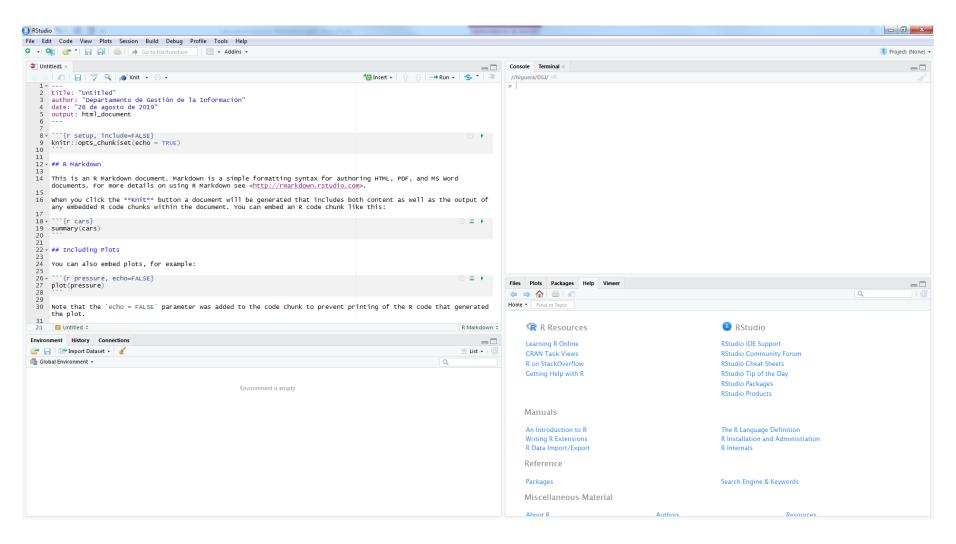


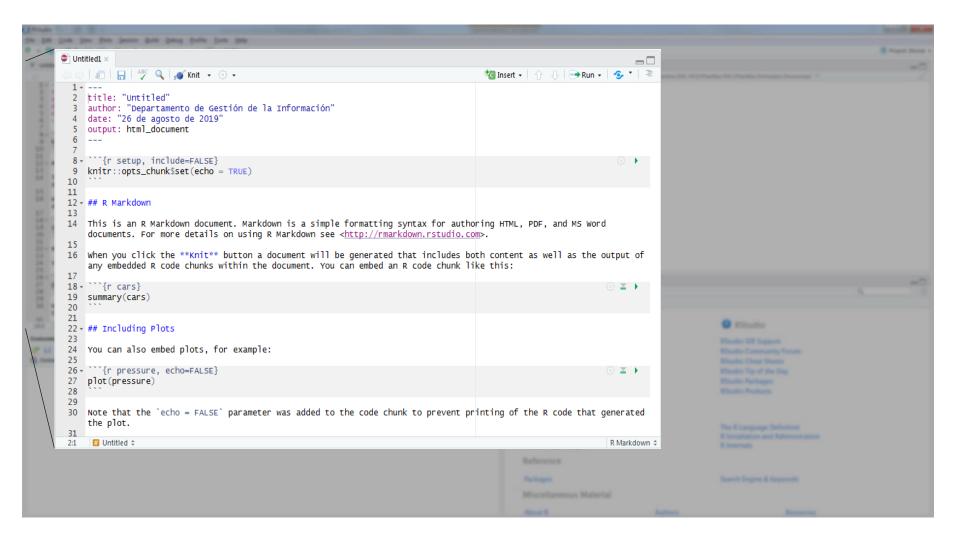
Entorno para la creación de reportes/documentos reproducibles

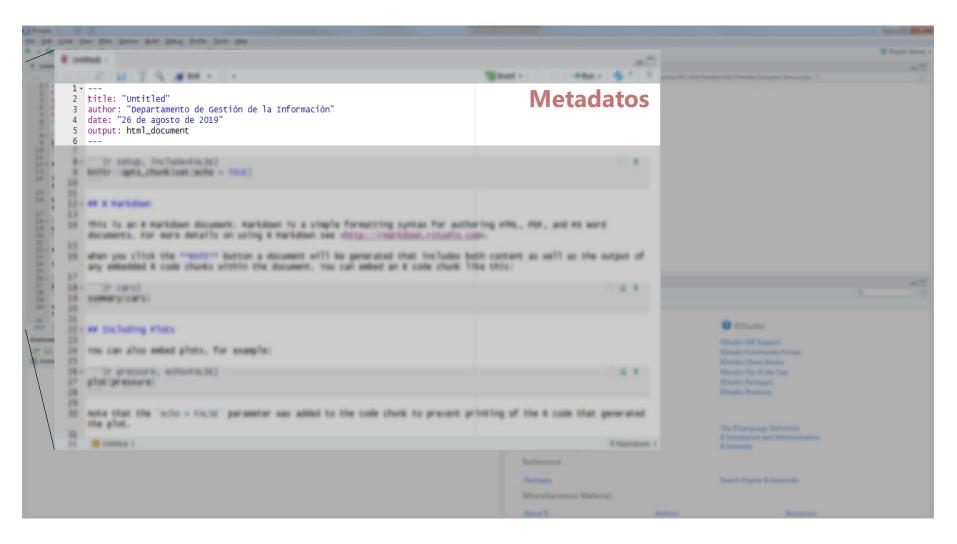
- Microsoft Word
- HTML
- PDF
- Power Point
- Y más...

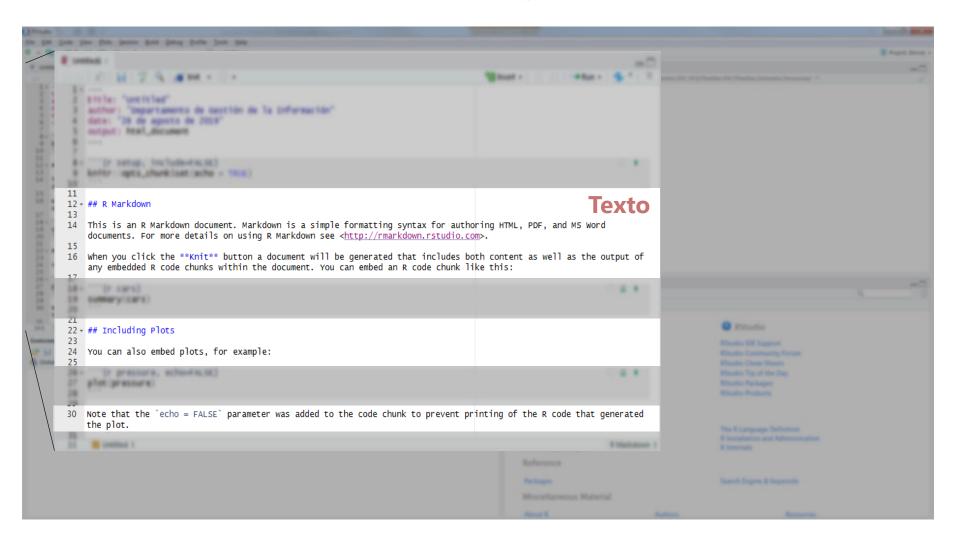


https://rmarkdown.rstudio.com











```
50 - # 1. Sobre la información presentada [/1]
52 - ###### A continuación, se presenta una ejemplo de minuta con diferentes indicadores asociados a la **'r reg'**. Los datos son simulados y no
    deben ser considerados en ningún caso como información oficial de la SMA. Este reporte fue generado el **`r format(Sys.Date(), "%d/%m/%v")`**.
53
    [A1]: **UF**: Unidad Fiscalizable / **RCA**: Resolución de Calificación Ambiental / **IA**: Inspección Ambiental / **PPDA**: Plan de Prevensión y
54
    Descontaminación Ambiental / **RSMA**: Resolución de normas e instrucciones de carácter de la Superintendencia de Medio ambiente / **PC/PDC**:
    Programa de Cumplimiento / **NC**: Norma de Contaminación / **LEY**: Ley Ambiental / **REG**: Reglamento / **NE**: Norma de Emisión / **MP**:
    Medida Provisional / **FDC**: Formulación de Cargos / **MM**: Millones
57 - # 2. Unidades Fiscalizables
58
59
    - **`r prettyNum(nrow(UF_reg), big.mark = ".", decimal.mark = ",")`** de las **`r prettyNum(nrow(Resumen_UnidadFiscalizable), big.mark = ".",
    decimal.mark = ",") ** Unidades Fiscalizables (UF) a nivel nacional corresponden a la ** r reg ** ( r pasteO(Porcentaje_uf_reg, "%") ).
   - Estas se concentran mayoritariamente en la categoría **`r resumen_cate[1,1]`** (`r resumen_cate[1,2]` UFs), seguida por la **`r
    resumen_cate[2,1]`** (`r resumen_cate[2,2]` UFs)
61
62 - ## 2.1 Localización UFs en la 'r reg'
63
    ```{r Mapa UFs, out.width="400px", fig.align="center"}
65 knitr::include_graphics("Mapa/mapa.png")
```

## 6.1 Formulaciones de cargo (FDC) &



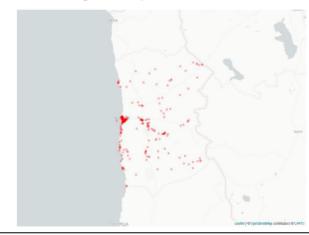
#### 1. Sobre la información presentada <sup>1</sup>

A continuación, se presenta una ejemplo de minuta con diferentes indicadores asociados a la Región de Tarapacá. Los datos son simulados y no deben ser considerados en ningún caso como información oficial de la SMA. Este reporte fue generado el 07/06/2020.

#### 2. Unidades Fiscalizables

- 362 de las 15.773 Unidades Fiscalizables (UF) a nivel nacional corresponden a la Región de Tarapacá (2.3%).
- Estas se concentran mayoritariamente en la categoría Equipamiento (86 UFs), seguida por la Minería (42 UFs)

#### 2.1 Localización UFs en la Región de Tarapacá

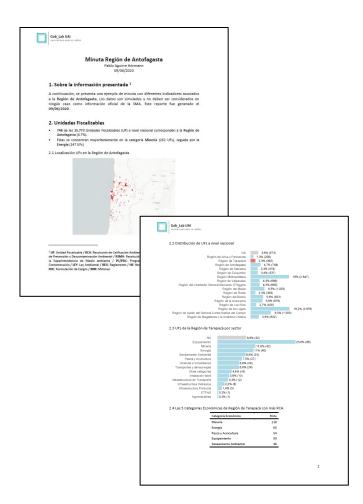


## Ejemplo

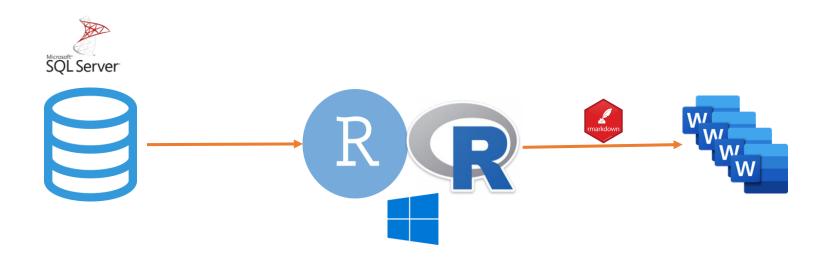
### Minutas de información general por región

- Definir parámetros
- Cargar librerías
- Importar datos
- Realizar modificaciones a los datos
- Generar cuerpo del reporte Texto

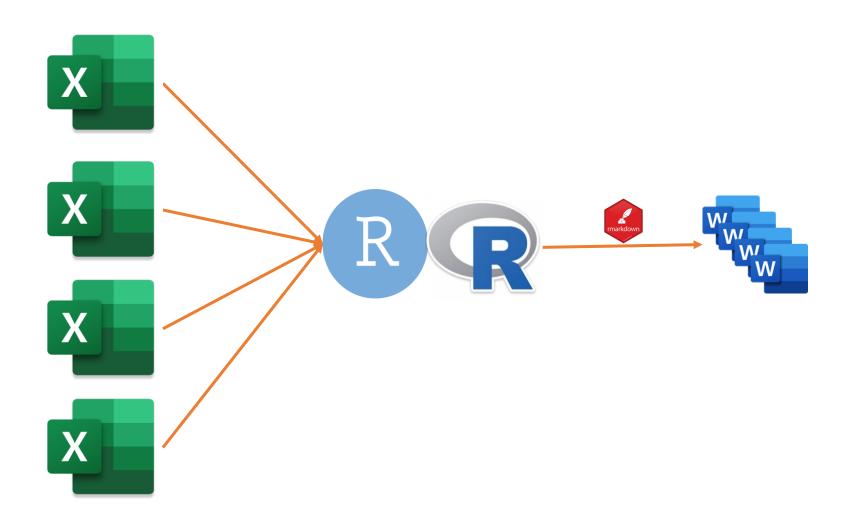
Tablas y gráficos



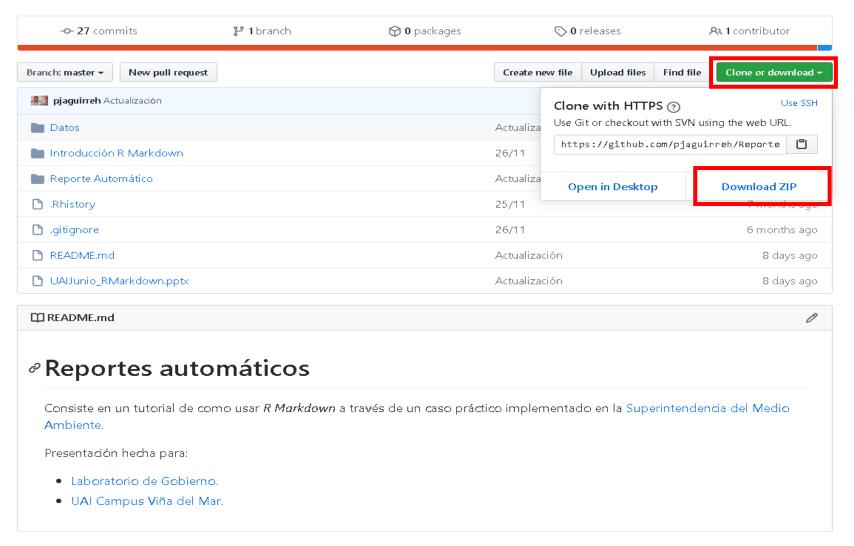
## Idea general de pipeline



## En este ejemplo



## Si tienes instalado R/RStudio GitHub



https://github.com/pjaguirreh/Reportes-automaticos

## Demostración



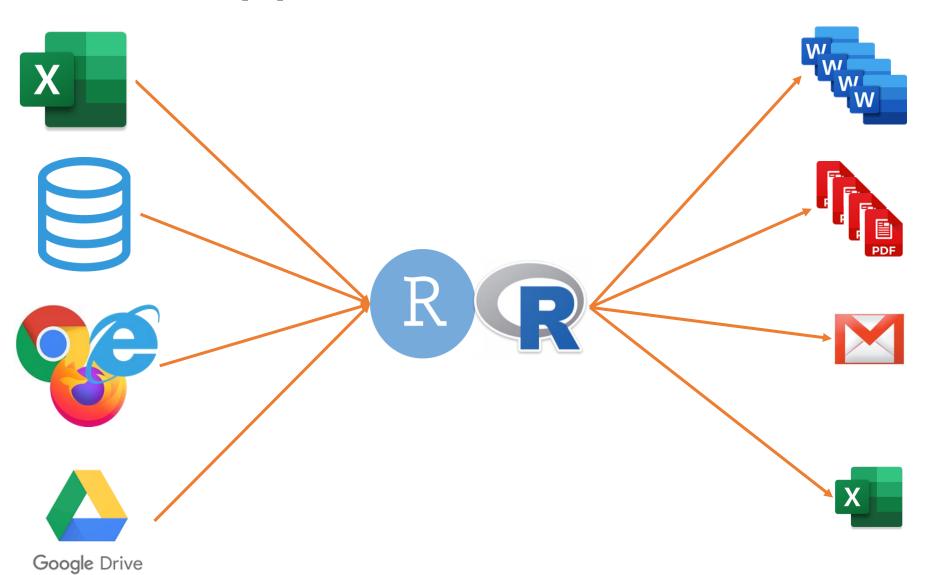




## Idea general de pipeline



## Posibles pipelines



## Dashboards en R - Shiny

#### Visualizador Covid-19 Chile ☼



▲ Descargar datos

#### Cifras generales 💝

Última actualización: 29 de mayo

Casos activos: 51096

Casos nuevos: 3695

Casos totales a la fecha: 90638

Casos Región Metropolitana: 72910

Región con más casos (exceptuando la RM): Valparaíso, con 2842 casos

Región con menos casos: Aysén, con 16 casos

Comuna con más casos: Santiago, en Metropolitana, con 4618 casos

Número de fallecidos totales: 944

Pacientes hospitalizados en UCI: 1350 personas

Exámenes PCR realizados en total: 485950 exámenes desde el 9 de abril



Todos los datos son obtenidos en tiempo real desde el repositorio de datos de la *Mesa de Datos COVID-19* del Ministerio de Ciencia, Tecnología, Conocimiento e Innovación.

Pueden presentarse desfases entre la información regional y comunal por las fechas de publicación de informes.

Aplicación desarrollada por DATA UC

Facultad de Matemáticas

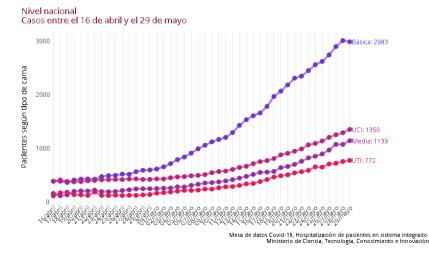
Pontificia Universidad Católica de Chile

Esta aplicación presenta de forma gráfica la información relativa a la evolución de la pandemia del Coronavirus COVID-19 en Chile, a partir del repositorio de la Mesa de Datos liderada por el Ministerio de Ciencia, Tecnología, Conocimiento e Innovación. Seleccione una sección para explorar los datos.



#### Pacientes hospitalizados

A continuación, se entrega una visión general del sistema de salud chileno a partir de datos diarios sobre la cantidad de pacientes en camas Básicas, Media, UCI o UTI, reportados por la Unidad de Gestión de Camas Críticas del Ministerio de Salud.



#### Hospitalizados totales por grupo de edad y género

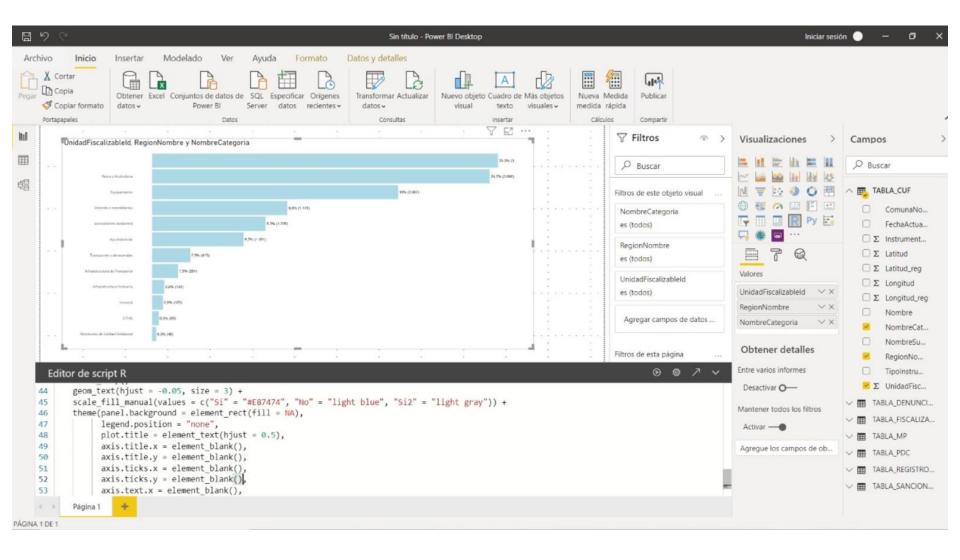
Las dos gráficas siguientes presentan la cantidad total de personas hospitalizadas desagregadas por género y tramo de edad, entregando una descripción sociodemográfica de los efectos del Covid-19 en la población.

https://shiny.rstudio.com/

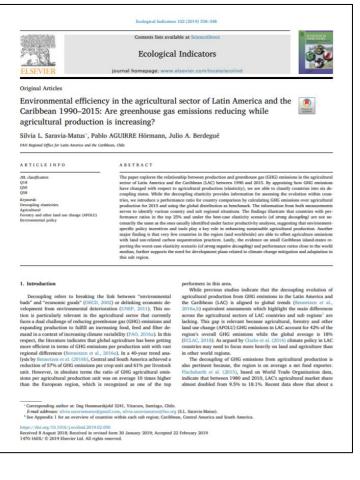
https://coronavirus.mat.uc.cl/

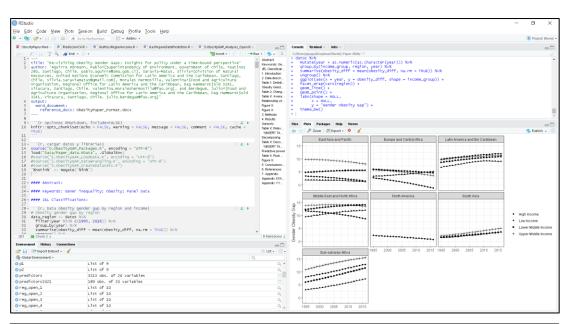
## Escribir código R en Power Bl

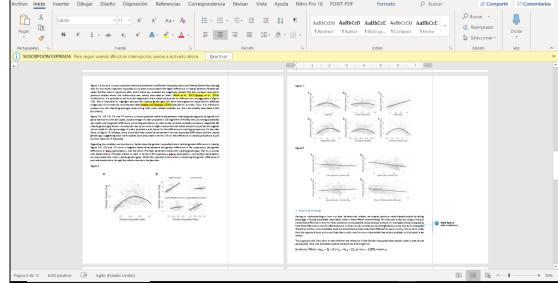




### Artículos académicos

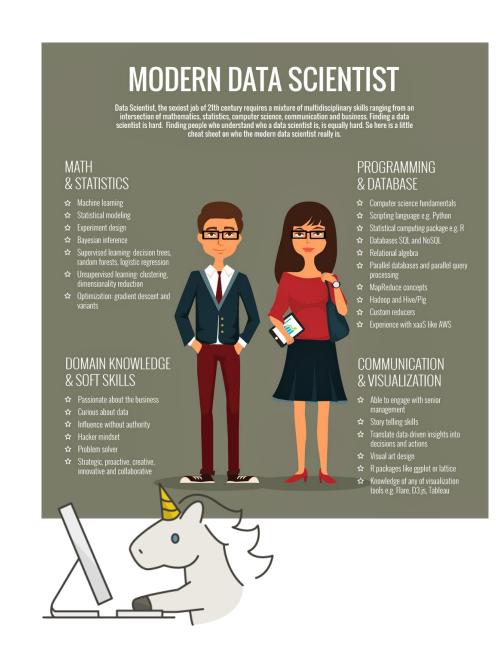






## Trabajo en equipo

- Equipos BI/Analytics
- Desarrolladores
- Infraestructura/redes
- Áreas de negocio/clientes
- Jefaturas
- Y más...



# Automatización de reportes (y un poco más) con R



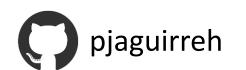


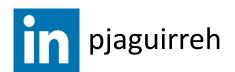




pjaguirreh@gmail.com







## Inteligencia Ambiental 2020: La Estrategia Tecnológica que refuerza la SMA para incentivar el cumplimiento ambiental

La iniciativa de la Superintendencia del Medioambiente -que busca incentivar el cumplimiento y la detección temprana de desviaciones- incluye desde reporte electrónico, análisis automatizados, uso de imágenes satelitales, hasta el envío de notificaciones masivas a empresas reguladas respecto de su desempeño ambiental.

**Santiago, 4 de junio de 2020.-** Con el fin de potenciar las capacidades instaladas y consolidar la plataforma electrónica de la SMA, mediante el uso de herramientas analíticas y automatización de procesos, el Departamento de Gestión de la Información lanzó su estrategia "Inteligencia Ambiental 2020."

La iniciativa, que fomenta la detección temprana de irregularidades con respecto a la normativa ambiental permite que desde fines del mes pasado la Superintendencia del Medio Ambiente diera inicio al envío de notificaciones masivas sobre el desempeño ambiental de fuentes afectas al Decreto Supremo N°90/2000 (norma de emisión de residuos líquidos) y a Centros de Engorda de Salmones (CES) que cuentan con Resolución de Calificación Ambiental (RCA).

En el primero de los casos se envió un reporte vía correo electrónico a 727 fuentes, o que equivale a 690 Unidades Fiscalizables afectas a la norma de emisión de residuos líquidos, representando el 100% del universo total de regulados.

Los reportes enviados contienen un análisis de los autocontroles realizados por las empresas durante el último mes, a la vez que entregan información sobre distintos aspectos regulados en sus autorizaciones, como lo son reporte, frecuencia y parámetros, entre otros.