

R-Ladies Puebla + csv,conf,v8

Intro a R para analizar CSV

Orden del Día



Hora	Actividad
12:00-12:05	Que es R-Ladies
12:05-12:10	RevisarCodigo de Conducta
12:10-12:15	Anuncios de R-Ladies Puebla
12:15-12:30	Instalación de programas
12:30-12:45	Introducción a R y Limpieza de Datos
12:45-13:00	Descanso
13:00-14:30	Continuación de Limpieza de Datos
14:30-15:00	Generación de graficos con ggplot2

R-Ladies Código de Conducta



R-Ladies está dedicado a proveer una experiencia libre de acoso para todas y todos sus participantes. No toleramos ninguna forma de acoso en cualquiera de sus formas.

Este código de conducta aplica a todos los espacios de R-Ladies, incluyendo meetups, Twitter, Slack, listas de correo, tanto en sesiones en línea como fuera de ella. Cualquier persona que viole este código de conducta podrá ser sancionada o expulsada de cualquiera de estos espacios a discreción del comité organizador.

Algunos espacios de R-Ladies pueden tener normas adicionales, las cuales se harán de claro conocimiento a las y los participantes de dichos espacios. Las y los participantes son responsables de conocer y atender las normas mencionadas.

Si estás siendo acosadx por un o una miembrx/invitadx/participante/ en R-Ladies Puebla, o notas que alguien más está siendo acosada/o o tienes alguna otra inquietud, comunícate con el comité organizador y realiza un reporte mediante el correo puebla@rladies.org

Anuncios de R-Ladies Puebla



- Estamos buscando organizadoras
- Si estas interesado pueden enviar un correo a puebla@rladies.org o en Twitter [@rladiespuebla](https://twitter.com/rladiespuebla)
- R-Ladies Global está llevando a cabo una discusión comunitaria sobre un posible cambio de nombre y queremos recopilar la mayor cantidad de opiniones posibles de la comunidad. Si prefieren, pueden compartir sus comentarios de forma anónima. Pueden encontrar la discusión en [GitHub](#).

Que es R-Ladies



R-Ladies es una organización internacional cuya misión es promover la diversidad de género en la comunidad R, a través de reuniones presenciales, virtuales y mentorías en un espacio amigable.

R-Ladies Puebla fue fundada en 2018 y tenemos 2,465 integrantes

R-Ladies Stats:

- Más de 100,000 integrantes
- 233 capítulos
- 63 países

Que Necesitas para Este Taller



- Programa: [R y RStudio Instalada](#).
- Materiales: [Github repo](#)

Instalando y Cargando paquetes

“Instalar” es cuando descargas el paquete en tu computadora.

```
1 # Display and execute the install.packages command
2 install.packages("tidyverse")
```

The downloaded binary packages are in
/var/folders/tm/ybqzrprj27jlp7jr0bwv2_0h0000gn/T//RtmpaI8kW4/downloaded_packages

“Cargar” es cuando le indicas a R que utilice el paquete.

```
1 library(tidyverse)
```

- Solo necesitas instalar una vez
- Necesitas cargar paquetes cada vez

Cargando nuestro dataset de hoy

```
1 data <-  
2   jsonlite::fromJSON('https://raw.githubusercontent.com/rladies/meetup_archive')
```


Funciones Principales

Función	Utilidad
<code> ></code>	pasar datos de una función a la siguiente
<code>mutate()</code>	crear, transformar y redefinir columnas
<code>select()</code>	mantener, eliminar, seleccionar o renombrar columnas
<code>arrange()</code>	ordenar las filas
<code>filter()</code>	filtrar filas de un dataframe basado en condiciones específicas.
<code>rename()</code>	cambiar el nombre de las columnas

Funciones Principales

Muchas de estas funciones pertenecen al paquete `dplyr`, que proporciona funciones “verbales” para resolver los retos de la manipulación de datos (el nombre hace una referencia a unos alicates - plier - de dataframes). `dplyr` forma parte de la familia de paquetes de R `tidyverse` (que también incluye `ggplot2`, `tidyr`, `stringr`, `tibble`, `purrr`, y `magrittr` entre otros).

Limpieza de datos

```
1 library(tidyverse)
2 rladies_chapters <- data |>
3   select(1,2,3,7,8) |> #names
4   rename(chapter=group_urlname) |>
5   mutate(location=ifelse(location=="Online event","online","inperson"),
6           title=sub(".*-- ","",title),
7           #title=sub(".*- ","",title),
8           title=gsub("\\s*\\([^\s\\)]+\\s)","",title)) |> #head
9   filter(!str_detect(title,regex("canceled|cancelled",ignore_case=T)),
10          !chapter%in%c("RLadiesJeddah","muhq_deleted@4633@rladies-ushuaia",
11                        "muhq_deleted@9919@notopic@508502","notopic@544550")) |>
12   arrange(desc(date)) |>
13   filter(year(date)<2024) |>
14   mutate(year=year(date))
```

```
1 head(rladies_chapters)
```

```
      id      chapter
1 298000525    rladies-rtp
2 297631808 rladies-ljubljana
3 297750427    rladies-rome
4 297305466    rladies-taipei
5 297335433    rladies-urmia
6 297762822    rladies-rome

                                title      date
1      [ONLINE] R-Ladies RTP Informal Hangout & Networking 2023-12-28
2                                Delavnica: PONOVLJIVA ZNANOST? 2023-12-20
3      R-Ladies Rome - Meet and Greet, 20 Dec 2023, 05:30 PM 2023-12-20
4                                跟著 AI 叮叮噹! 2023-12-18
5 Trend, Seasonality. Cycling, and outliers in Time Series Analysing 2023-12-15
6      Data Visualization Practice Session - in Italian 2023-12-15

location year
1  online 2023
2 inperson 2023
3 inperson 2023
```

Data Wrangling

```
1 dat <- rladies_chapters |> #View
2   group_by(chapter, year) |>
3   count(chapter) |>
4   arrange(-year)
5
6   View(dat)
```

Data Wrangling Continued

```
1 # Viendo los top 50 capitulos de R-Ladies
2 top_50 <- dat |>
3   ungroup() |>
4   arrange(-year) |>
5   group_by(chapter) |>
6   summarize(tot=mean(n)) |>
7   arrange(-tot) |>
8   slice(1:50)
9
10 View(top_50)
```

Data Wrangling 3

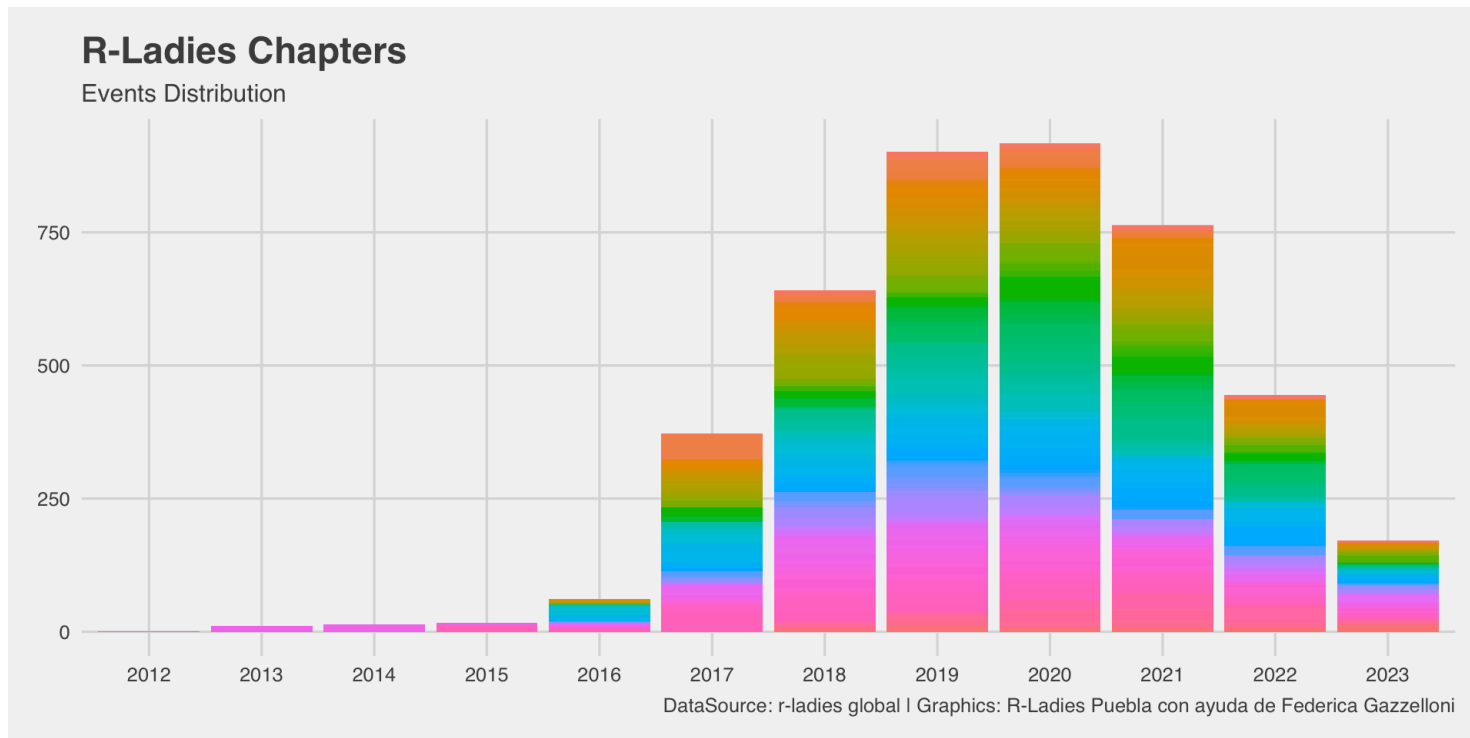
```
1 days_week <- rladies_chapters |>
2   select(date, location, year) |>
3   mutate(date = lubridate::wday(date, label = TRUE, week_start = 1)) |>
4   count(date, location, year) |>
5   mutate(location = case_when(
6     location == "inperson" ~ "In person",
7     location == "online" ~ "Online"
8   ))
```

Empezando con ggplot2

ggplot2 es el paquete de R más popular para la visualización de datos. Su función `ggplot()` es el núcleo de este paquete, y todo este enfoque se conoce coloquialmente como “ggplot”, con las figuras resultantes a veces llamadas afectuosamente “ggplots”. El “gg” en estos nombres se refiere a la “gramática de los gráficos” utilizada para construir las figuras. `ggplot2` se beneficia de una amplia variedad de paquetes de R complementarios que mejoran aún más su funcionalidad.

Primera Grafica

```
1 dat |>
2   ggplot(aes(x=factor(year),y=n,fill=chapter))+
3   geom_col(show.legend = F)+
4   ggthemes::theme_fivethirtyeight()+
5   labs(title="R-Ladies Chapters",
6        subtitle="Events Distribution",
7        x="año",y="n. de eventos",
8        caption="DataSource: r-ladies global | Graphics: R-Ladies Puebla con ay
```



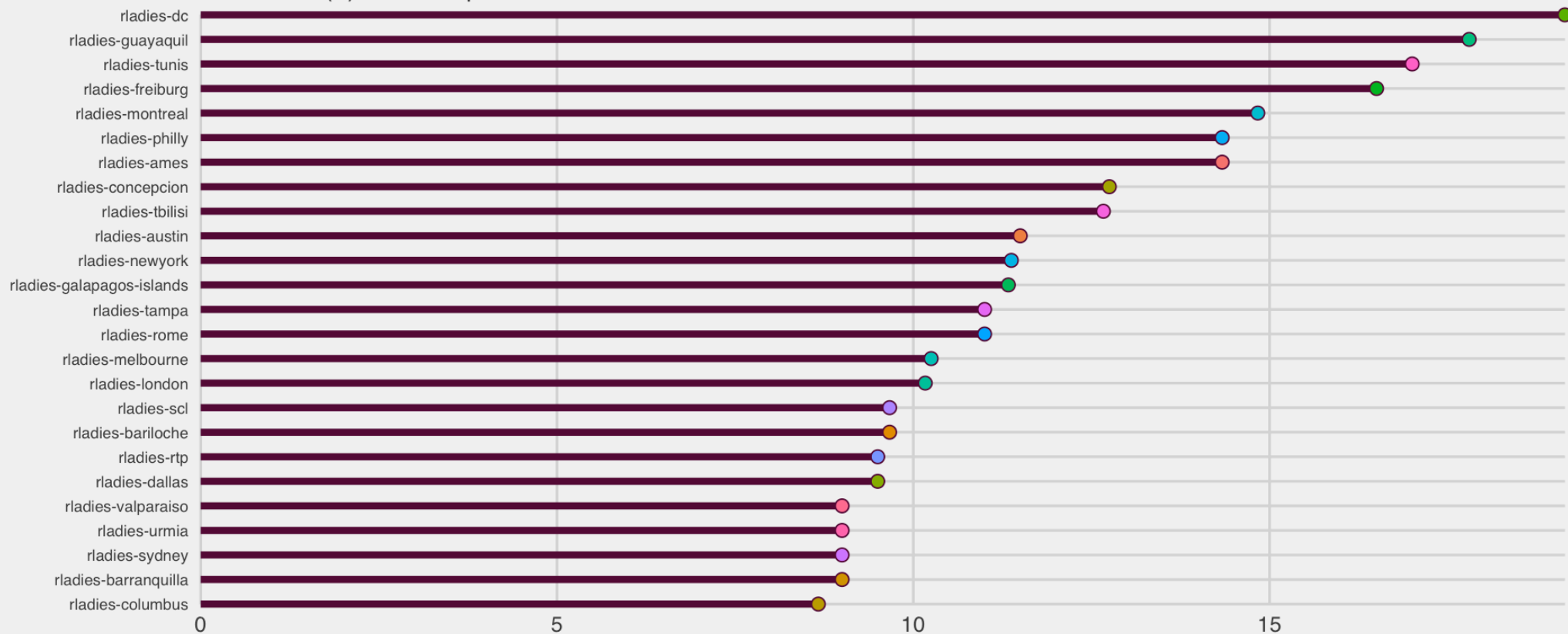
Segundo Grafica Limpiza de Datos

```
1 top25 <- dat|>
2   #ungroup()|>
3   #arrange(-year)|>
4   group_by(chapter)|>
5   summarize(tot=mean(n))|>
6   arrange(-tot)|>
7   slice(1:25)
```

Segunda Grafico Visualización

Top 25 Capítulos de R-Ladies

Promedio (n) eventos por año



DataSource: R-Ladies Global | Graphics: R-ladies Puebla con ayuda de Federica Gazzelloni

Tercera grafica codigo

```
1 #instalar y cargar packages
2 install.packages("gghighlight")
```

The downloaded binary packages are in

/var/folders/tm/ybqzrprj27jlp7jr0bwv2_0h0000gn/T//RtmpaI8kW4/downloaded_packages

```
1 library(gghighlight)
2
3 # practicar cargando datos otra vez
4 data_new <- readr::read_csv('https://raw.githubusercontent.com/rfordatascience
5
6 # Celdas explícitas y crear columna para contar
7 data_new <- data_new |>
8   complete(chapter, year, location) |>
9   mutate(event = ifelse(is.na(date), 0, 1))
10
11
12 # Echa un vistazo a los capítulos con el mayor número de eventos
13 dat_agg_chap <- data_new |>
14   group_by(chapter) |>
15   summarise(total=sum(event))
16
17 # Crear un vector de los 10 capítulos principales por número de eventos
18
19 topten <- dat_agg_chap |>
20   top_n(10)
```

R-Ladies theme for Quarto Presentations. Code available on GitHub.

Tercera Figura Visualización

Cantidad de eventos de R-Ladies para los capítulos con el mayor número de eventos

Las líneas continuas representan eventos presenciales, mientras que las líneas discontinuas representan eventos en línea

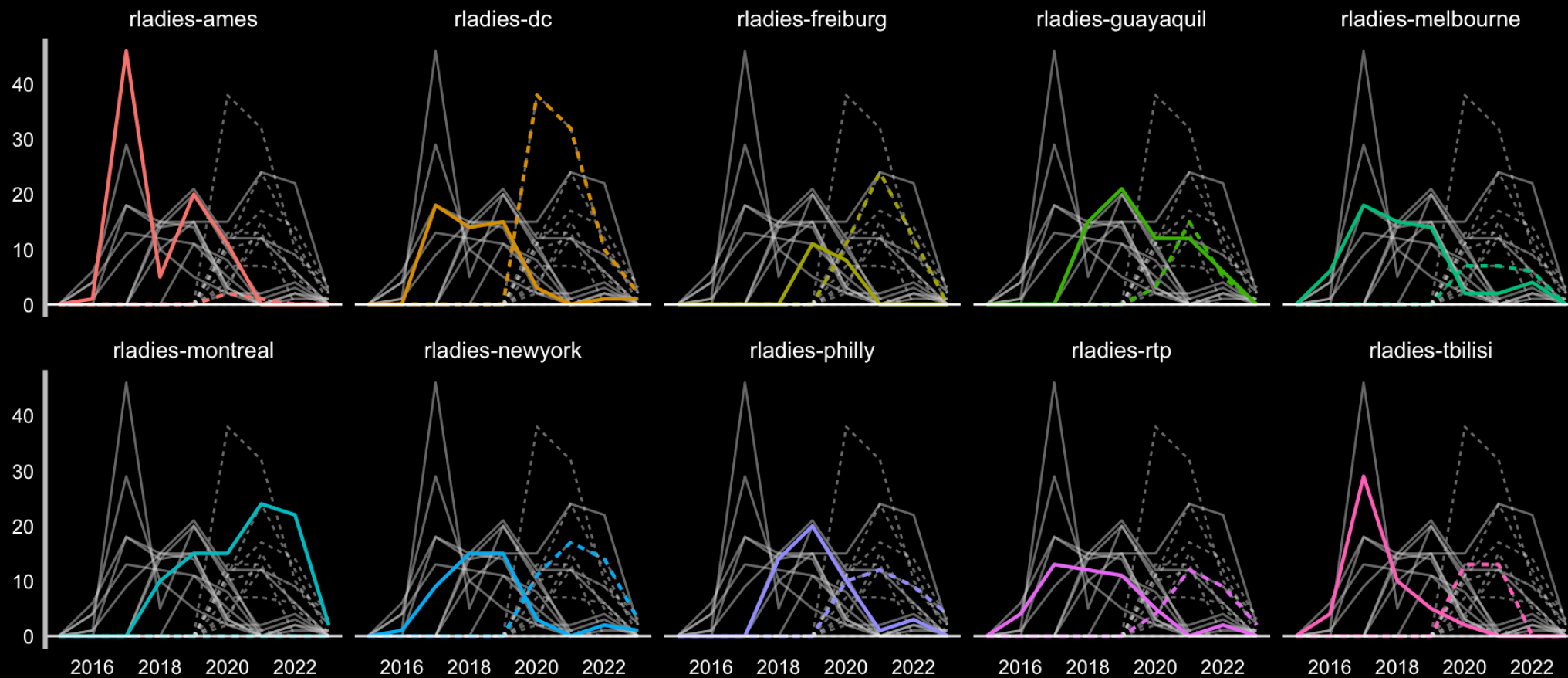


Tabla Datos

```
1 library(dplyr)
2 library(knitr)
3 library(kableExtra)
4 library(lubridate)
5
6 puebla_members_df <- read.csv('data/rladies_puebla_members.csv')
7
8 # Convert the date column to Date type
9 puebla_members_df$date <- as.Date(puebla_members_df$date)
10
11 # Create a summary table of the total and active members per month
12 summary_table <- puebla_members_df |>
13   mutate(month = floor_date(date, "month")) |>
14   group_by(month) |>
15   summarise(
16     total_members = mean(total),
17     active_members = mean(active)
18   ) |>
19   arrange(month)
```

Tabla Visualización

Monthly Summary of Total and Active Members in Puebla

Month	Average Total Members	Average Active Members
2021-09-01	2132.045	277.000000
2021-10-01	2214.722	275.944444
2021-12-01	2222.400	8.366667
2022-01-01	2224.200	8.200000
2022-02-01	2227.370	9.296296
2022-03-01	2249.226	83.258065
2022-04-01	2355.233	228.433333
2022-05-01	2416.258	104.870968
2022-06-01	2417.821	14.250000
2022-07-01	2416.774	7.258064
2022-08-01	2417.161	4.290323
2022-09-01	2416.000	2.233333
2022-10-01	2417.613	7.548387
2022-11-01	2423.333	7.300000
2022-12-01	2422.548	2.419355
2023-01-01	2422.161	1.290323
2023-02-01	2424.393	4.250000
2023-03-01	2426.516	6.645161
2023-04-01	2427.333	5.600000
2023-05-01	2429.097	4.451613
2023-06-01	2429.071	5.785714
2023-07-01	2433.067	4.333333
2023-08-01	2437.258	4.774193
2023-09-01	2439.133	4.200000
2023-10-01	2442.774	5.064516
2023-11-01	2443.833	2.233333
2023-12-01	2447.839	3.903226
2024-01-01	2449.611	3.333333

Preguntas??

Agradecimiento

csv, conf, v8

Federica Gazzelloni (github)[<https://github.com/Fgazzelloni>]

csv,conf,v8

Jueves 10:00 am Explorando Desafíos y Fortalezas: R-Ladies
Puebla y la Ciencia de Datos con Perspectiva de Género