



Tecnológico
de Monterrey

01. Presentación

Módulo 5. Inteligencia Artificial

Jorge
Juvenal
Campos Ferreira

juvenal.campos@tec.mx

Programa de la clase

- Presentación
- Revisión del temario
- Diagnóstico programación
- Verificación de instalación de R
- Revisión mecanismos evaluación
- Recomendaciones uso IA
- Uso de LLMs

Sobre mí



M.C. JORGE JUVENAL CAMPOS FERREIRA.

- * **Analista de datos, México, ¿Cómo vamos? Y Fundación Novagob México**
- * **Columnista en Atiempo.TV Coahuila**

Educación Formal:



Licenciatura:

Ingeniería en Irrigación por la Universidad Autónoma Chapingo. (2009-2014).



Maestría:

Maestría en Economía por El Colegio de México (2016-2018).

Contacto:



GitHub: JuveCampos



LinkedIn: Jorge Juvenal Campos Ferreira



Twitter: @JuvenalCamposF



IG: juvenalcampos.dataviz

Experiencia profesional

- **Analista de datos.**
 - CIDE - Laboratorio Nacional de Políticas Públicas.
 - México ¿Cómo vamos?
 - Fundación Novagob México
 - Clave Igualdad
- **Profesor**
 - Periodismo de datos - Maestría en Periodismo del CIDE
 - Tableros en R/Shiny - Datacrunchers
- **Periodista**
 - Columna semanal en Atiempo.TV
 - Escritor y colaborador en Nexos y Animal Político

Instalación R y Python

- Vamos a verificar que tengan R y Python instalado.



Instalar R

<https://cran.r-project.org/>



Instalar RStudio

[https://posit-co.translate.goog/download/rstudio-desktop/?
_x_tr_sl=en&_x_tr_tl=es&_x_tr_hl=es&_x_tr_pto=tc](https://posit-co.translate.goog/download/rstudio-desktop/?_x_tr_sl=en&_x_tr_tl=es&_x_tr_hl=es&_x_tr_pto=tc)

El curso se va a dar principalmente en R y RStudio, por lo que sí es importante que verifiquen que el programa esté instalado y funcionando.

Test diagnóstico

Cuestionario para saber como se encuentra el grupo en materia de IA y manejo de datos.



[https://docs.google.com/forms/d/e/
1FAIpQLSegYnFGwp6yNx3T95WIPJhUIX4c968IkeAfqYU0SYg1VC8udg/viewform](https://docs.google.com/forms/d/e/1FAIpQLSegYnFGwp6yNx3T95WIPJhUIX4c968IkeAfqYU0SYg1VC8udg/viewform)

Sobre ustedes

- ¿Cómo se llaman?
- ¿De donde son?
- ¿Por que están en esta carrera?
- ¿Que tanto saben de Ciencia de Datos?
- ¿Qué herramientas saben usar?
- ¿Qué esperan de este curso?



Propósito del curso

El curso busca que el estudiante:

Dotar a los y las estudiantes con las herramientas necesarias para el manejo de datos y desarrollo de simulaciones, así como reforzar los conocimientos adquiridos y profundizar en el desarrollo de las competencias mediante la experimentación y la práctica. Los temas se proponen a partir de efectos prácticos.

Lo que yo busco



1. Que se sientan más seguros al programar
2. Que estén más preparados para los efectos de la IA en los próximos años
3. Que puedan resolver problemas prácticos de su ámbito profesional.

**¿Por qué economía?
¿A que se dedica un economista?
¿Qué labores hace en su día a día?**

1. **Introducción a la IA y a los grandes modelos de lenguaje**
 - a. Peligros de su uso y abuso de la IA
 - b. Beneficios en el uso de la IA
 - c. Fundamentos de los LLMs
 - d. Mejora de prompts (introducción al *prompt-engineering*)
 - e. Herramientas de generación de texto
 - f. Herramientas de *vibe coding*
 - g. Otras herramientas
 - h. Casos de uso

2. **Programación en R**
 - a. Instalación de R
 - b. Uso de proyectos y preparación del entorno de trabajo
 - c. Objetos y funciones (que también son objetos)
 - d. Estructura de datos
 - e. Carga de archivos
 - f. Introducción al tidyverse
 - g. Manipulación básica de datos (filtrar, seleccionar columnas, ordenar, agrupar, resumir y encadenar funciones)
 - h. Estadística descriptiva
 - i. Bucles y funciones lapply
 - j. Trabajo con datos de SQL

3. Visualización de datos

- a. Visualización de datos con ggplot
- b. Visualización interactiva

4. Datos de México

- a. Fundamentos de encuestas e incertidumbre
- b. Datos del INEGI
- c. Encuesta Nacional de Ingresos y Gastos de los Hogares
- d. Encuesta Nacional de Ocupación y empleo
- e. Construcción de indicadores

5. Aplicaciones

a. Obtención de datos de internet

- i. Datos abiertos
- ii. Web scraping
- iii. Uso de APIs

b. Manejo de información geográfica y elaboración de mapas

6. Modelos de Machine Learning

- a. Regresión lineal
- b. Modelos de regresión
- c. Modelos de clasificación
- d. Pruebas de los modelos

Contenidos conceptuales

- 5.1 Tráfico vehicular en una ciudad.
- 5.2 Transmisión de un rumor.
- 5.3 Análisis del crimen y el terrorismo.
- 5.4 Cambio climático.
- 5.5 Competencia económica.

Evaluación

Criterios de evaluación:

El 5% de la evaluación correspondiente a este módulo se compondrá de lo siguiente:

==== Total ===

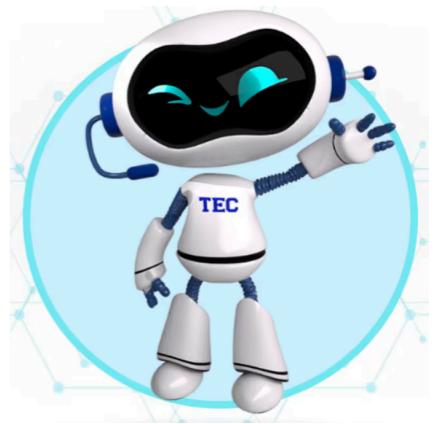
- * 2% (40%) Tarea al inicio del semestre
- * 3% (60%) Examen a mediados de semestre

Se realizará un examen, y una serie de tareas relativas a los temas que se revisarán.

Reglas de clase

- Mantener el **respeto** hacia compañeros y profesores
- **Participar** de forma ordenada y respetuosa
- Usar dispositivos electrónicos **solo cuando el profesor lo indique**
- **Realizar actividades** correspondientes al curso en desarrollo

Uso de la IA



✓ Se permite usar modelos de IA para el apoyo en la resolución de trabajos, siguiendo los **principios éticos** del uso de la IA en el Tec de Monterrey.

Respeto a la dignidad humana: No manipular ni influir indebidamente en personas.

No maleficencia: Evitar daños físicos, psicológicos, reputacionales o a la privacidad.

Promoción de la autonomía: Fomentar que las personas tomen decisiones informadas.

Equidad: Acceso inclusivo y beneficios compartidos.

Seguridad: Proteger datos y usar entornos seguros.

Veracidad: Contrastar y validar la información generada.

Explicabilidad y transparencia: Entender y declarar el uso de IA.

Responsabilidad: Evaluar consecuencias y actuar con reflexión.

Bienestar social y medioambiental: Uso para el bien común y la sostenibilidad.

Uso de la IA



Sugerencias:

1. Declarar siempre su uso en trabajos individuales y colaborativos
2. Emplear IA para potenciar el aprendizaje, resolver dudas y organizar ideas
3. Citar y referenciar cuando sea necesario (seguir formato APA para Chatbots)

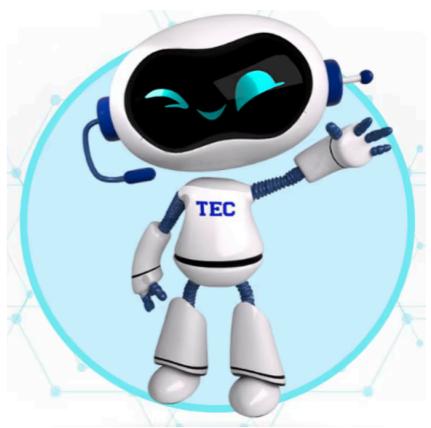
Estudiantes



Docentes



La IA no es excusa



Los errores de la IA que uses son tus errores.
Los errores de la IA que use alguien de tu equipo
son tus errores.
La deshonestidad que cometa tu uso de la IA
tendrá consecuencias sobre tí y tu equipo.

Estudiantes



Docentes



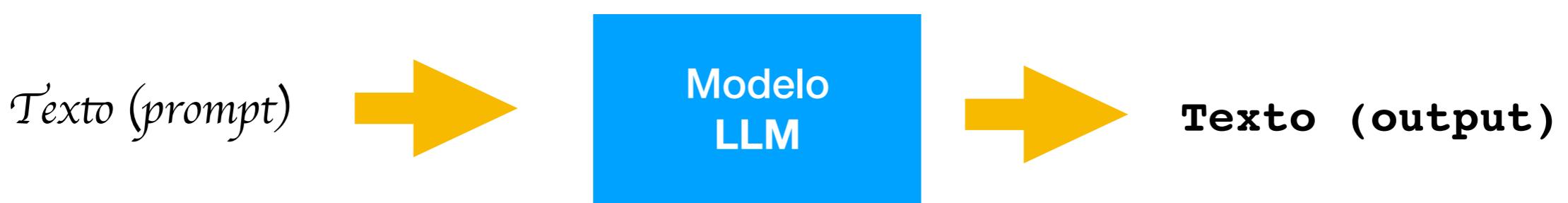
Diagnostico uso herramientas Ciencia de Datos

Modelos LLM

Introducción a los Modelos de Lenguaje

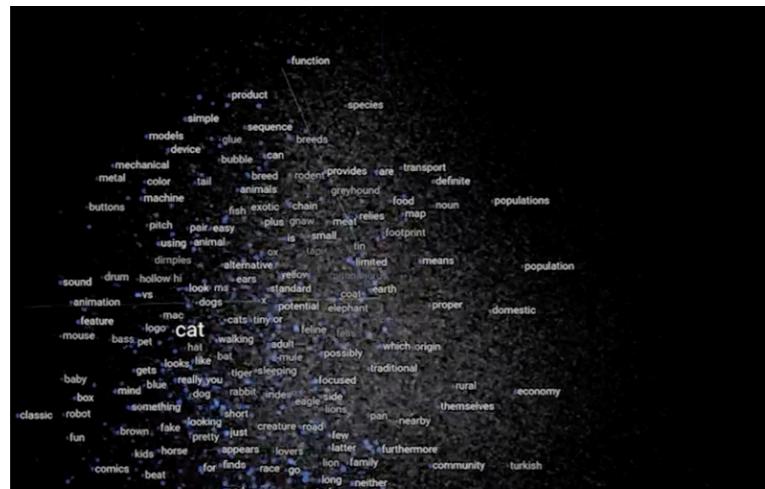
Un sistema de IA que **procesa y genera texto** en lenguaje natural, aprendiendo patrones a partir de **grandes** cantidades de datos.

Ejemplo: Predecir la palabra que sigue en una frase.

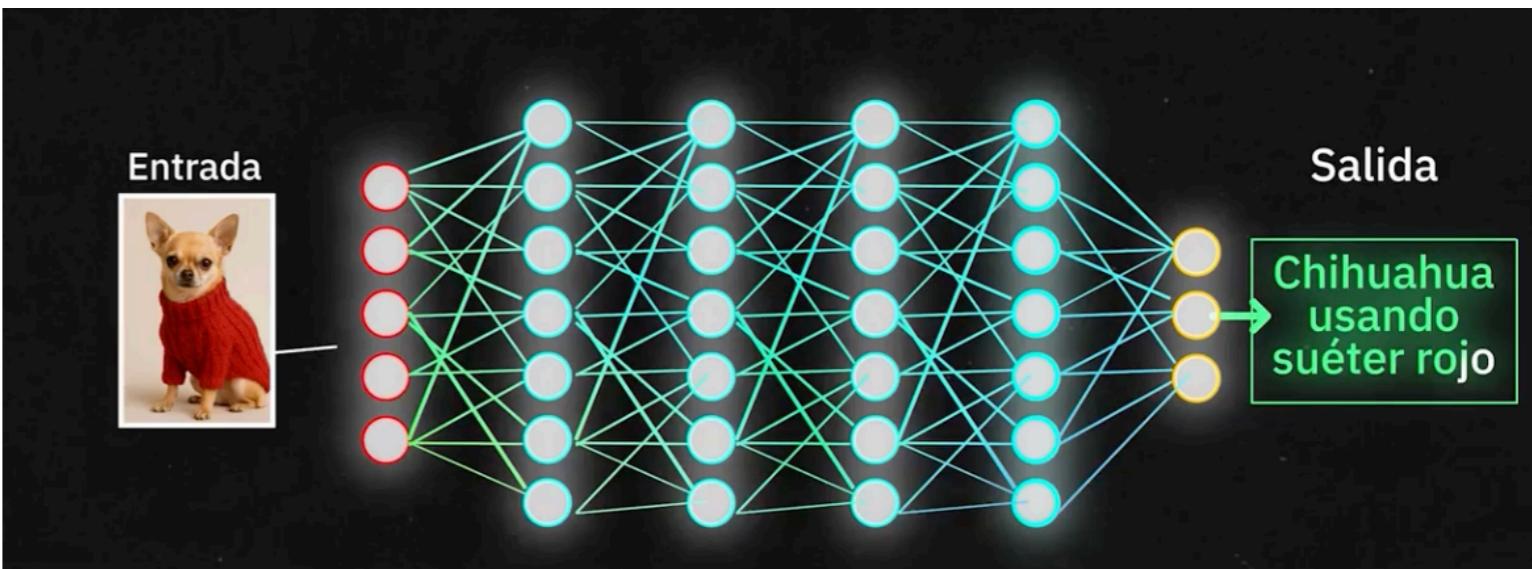
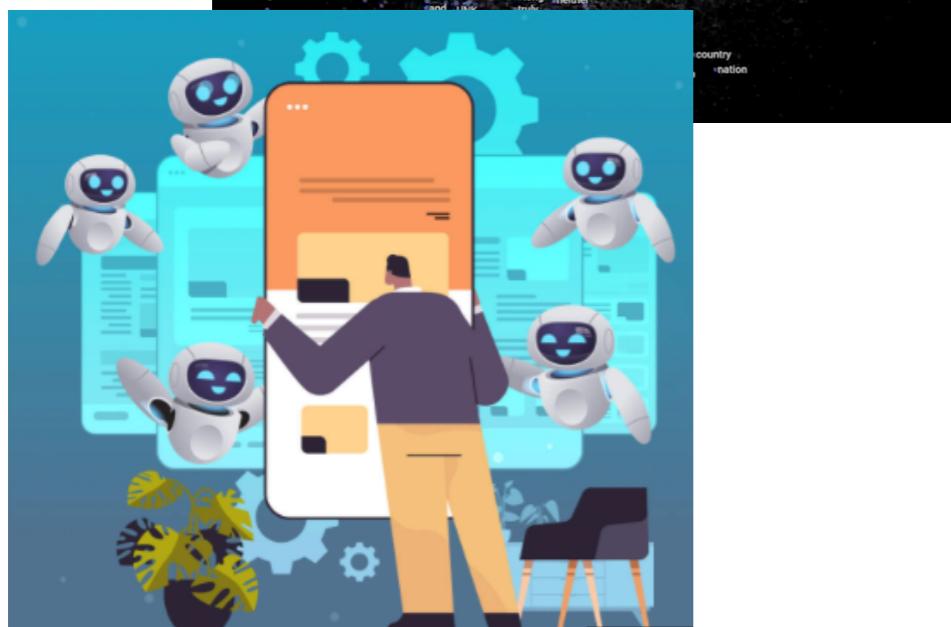


¿Cómo funciona un LLM?

1. Texto de entrada → **Tokenización** (palabras a números).
2. **Procesamiento** con redes neuronales (capas de **atención**).
3. Predicción de la **siguiente palabra** según el contexto.
4. Respuesta generada en texto natural.



$$A(Q,K,V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

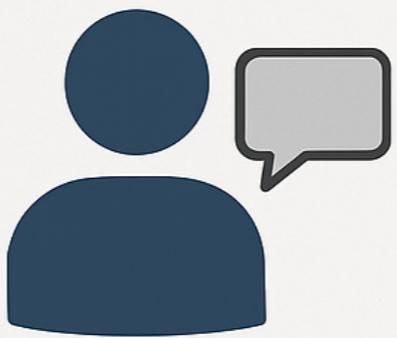


Usos comunes de los LLMs

Usos Comunes de los LLMs



Redacción y
edición de textos



Explicaciones
y tutorias



Resumen de
documentos



Simulación de
diálogos o roles



Resolución
de dudas



Generación
de código

**Las imágenes no se generan estrictamente en un modelo LLM, sino que se generan en otro tipo de modelos.*

¿Cómo funciona un LLM?

1.Revisar video:

<https://www.youtube.com/watch?v=awGfhmsN7Lc>



**Discutamos: ¿Es buena la IA? ¿Es mala?
¿Qué efectos nocivos puede tener su uso
y desarrollo?**

Fundamentos de Ingeniería de Prompts

Fundamentos de ingeniería de prompts

La “*ingeniería de prompts*” trata de incidir en las instrucciones que se le brindan a un LLM, de tal manera que:

- 1) Se transmita al modelo de manera inequívoca lo que se espera de él.
- 2) Lograr que el modelo genere respuestas consistentes y predecibles ante inputs similares
- 3) Obtener los mejores resultados con la menor cantidad de tokens posibles, optimizando tanto costos como tiempos de procesamiento.

La **calidad** de la respuesta depende directamente de la claridad, el contexto y la estructura de la instrucción.



Fundamentos de ingeniería de prompts

★ **Prompt:** Es la instrucción o pregunta que le damos al modelo para que genere una respuesta. La calidad del *prompt* influye directamente en la calidad de la respuesta.

Ingredientes mínimos de un buen prompt

Tarea: La actividad que queremos que realice el modelo.

Contexto: La información previa que queremos que el modelo tome en consideración al momento de procesar el prompt

Límites: Lo que el modelo debe o no debe incluir en su respuesta

Rol: El rol que debe asumir el modelo al momento de dar la respuesta.

Ejemplos de Prompts

Prompt sencillo

Realiza una investigación completa sobre el INFONAVIT de México. Incluye su historia, funcionamiento actual, noticias recientes y cualquier dataset disponible público. Proporciona información detallada y actualizada.



¿Cuál dará el mejor resultado?



Prompt más elaborado

```
<rol>
Actúa como un investigador especializado en instituciones de vivienda de México con experiencia en análisis de políticas públicas y manejo de datos gubernamentales.
</rol>

<enfoque>
Realizar una investigación exhaustiva y estructurada sobre el Instituto del Fondo Nacional de la Vivienda para los Trabajadores (INFONAVIT), abarcando múltiples dimensiones: histórica, institucional, operativa y de disponibilidad de datos.
</enfoque>

<contexto>
El INFONAVIT es una institución tripartita mexicana creada en 1972 que administra el fondo de vivienda de los trabajadores formales. Es crucial para el sector habitacional mexicano y maneja grandes volúmenes de datos sobre créditos, viviendas y trabajadores cotizantes. La investigación debe considerar su evolución, reformas recientes, impacto socioeconómico y transparencia en datos.
</contexto>

<tareas_específicas>
1. Investigar la historia institucional del INFONAVIT desde 1972 hasta la actualidad
2. Analizar noticias y desarrollos recientes (últimos 2 años)
3. Identificar y catalogar todos los datasets públicos disponibles
4. Examinar programas actuales y productos crediticios
5. Evaluar el marco regulatorio y reformas recientes
</tareas_específicas>

<limites>
- NO incluir información de otras instituciones de vivienda sin relación directa con INFONAVIT
- NO especular sobre datos no públicos o confidenciales
- NO incluir opiniones políticas partidistas
- Sí incluir únicamente fuentes verificables y oficiales
- Sí proporcionar enlaces directos a datasets cuando sea posible
- Sí mantener un enfoque analítico y objetivo
</limites>

<formato_salida>
Estructura la investigación en las siguientes secciones:
1. Resumen ejecutivo
2. Historia institucional
3. Marco legal y regulatorio actual
4. Programas y productos actuales
5. Noticias y desarrollos recientes (2023-2025)
6. Datasets disponibles (con URLs directas)
7. Análisis de transparencia y acceso a información
8. Conclusiones y recomendaciones para investigadores
</formato_salida>

<fuentes_prioritarias>
- Sitio oficial de INFONAVIT
- Portal de datos abiertos del gobierno mexicano
- INAI (Instituto Nacional de Transparencia)
- Informes anuales oficiales
- Boletines de prensa institucionales
- Bases de datos del IMSS relacionadas
</fuentes_prioritarias>
```

Ejercicio

Redacte un *prompt* para investigar un tema que le sea de interés.

Redáctelo tomando en cuenta **Tarea, Contexto, Límites y Rol**.

Utilice la sintaxis XML del ejemplo anterior (*envolver cada elemento en los símbolos “>” y “<”*). ¿Considera que redactar este prompt le representa alguna ventaja?

Prompts para otros modelos y casos de uso

Para videos:

- * Tarea
- * Historia: Descripción de lo que va a pasar en el video
- * Target audience: Descripción del público objetivo
- * Length: Tiempo de duración del video
- * Tono: Tono del diálogo del video
- * Formato: Formato del video (vertical, horizontal, 16:9)
- * Visuales: Descripción de los visuales del video, como el fondo, la iluminación, los personajes, la actitud, la mirada, etc.
- * B-Roll: Descripción del B-roll
- * Closing: Mensaje de cierre/Descripción del Cierre
- * Text and graphics: Mensaje a destacar, pantallas a mostrar, etc.



Para imágenes:

- * **Sujeto:** Sobre qué trata la imagen
- * **Acción:** ¿Qué está haciendo el sujeto?
- * Ambiente: ¿En donde se encuentra el sujeto?
- * **Estilo:** Estética general de la imagen (estilo anime, pintura antigua, hiperrealista)
- * **Iluminación:** Iluminación frontal, trasera, por fuego, luna, etc.
- * **Detalles clave:** Otros

Stacks de IA

Para empezar:

- * Una herramienta de chat general
- * Un buscador con fuentes
- * Una herramienta de creación
- * Una herramienta de automatización

Más experto:

- * Un copiloto de código
- * APIs de modelos
- * Agentes y orquestación
- * Automatización avanzada

LLMs para generar código

Modelos de generación de código

Los LLMs pueden escribir, corregir y optimizar código porque fueron entrenados no solo con lenguaje natural, sino también con **repositorios de código** y documentación técnica.

Aplicaciones:

- Autocompletar y escribir funciones enteras.
- Detectar y corregir errores (*debugging*).
- Explicar código existente.
- Convertir descripciones en código (prompt → script).
- Traducir entre lenguajes de programación.

Modelos de generación de código

- Hay muchas herramientas para generar código con IA. ChatGPT, Claude, Gemini o Grok funcionan bien
- También hay herramientas especializadas para mezclar IA y código, como Cursor, o CLIs (*Command Line Interface*), como Claude Code o Gemini CLI.
- Las herramientas especializadas en código facilitan el trabajo con este, y en algunos casos ejecutan y prueban el código (no hay que estar copiando y pegando, por ejemplo).



Modelos de generación de código

- Mi herramienta favorita (ahorita a agosto del 2025) es Claude Code.
- Es un CLI que permite acceder a carpetas de trabajo, leer y escribir archivos y también generar código.
- Claude Code ha tenido un buen desempeño al momento de trabajar con código de R, y particularmente en generar aplicaciones shiny básicas.



Claude Code

- Instrucciones de instalación: [https://www-anthropic-com.translate.goog/clause-code?
_x_tr_sl=en&_x_tr_tl=es&_x_tr_hl=es&_x_tr_pto=tc](https://www-anthropic-com.translate.goog/clause-code?_x_tr_sl=en&_x_tr_tl=es&_x_tr_hl=es&_x_tr_pto=tc)

Se necesita usar terminal.

Antes de la instalación, hay que instalar Node.js 18+



Ejercicio guiado

- * Instale Claude Code o Gemini en su computadora
- * Ejecútelo y corra un prompt que genere un archivo para generar una gráfica básica en *ggplot*.
- * Ejecute el código y verifique que funcione.

Ejercicio guiado 2

- * En el Github se encuentra el archivo “02_EJERCICIOS_PRACTICOS/Sesión_01/ejercicio_guiado_2.zip”. Descarguelo.
- * Descomprima el zip. La carpeta resultante tiene los datos de las proyecciones CONAPO al 2070. Inspeccione los datos.
- * Utilizando su agente de código y lo que vió de prompts, genere una aplicación shiny que le permite saber:
 - * ¿Qué estado tiene más población en un año determinado?
 - * ¿Qué años alcanza cada estado su población máxima?

Tareas que se hacen con datos

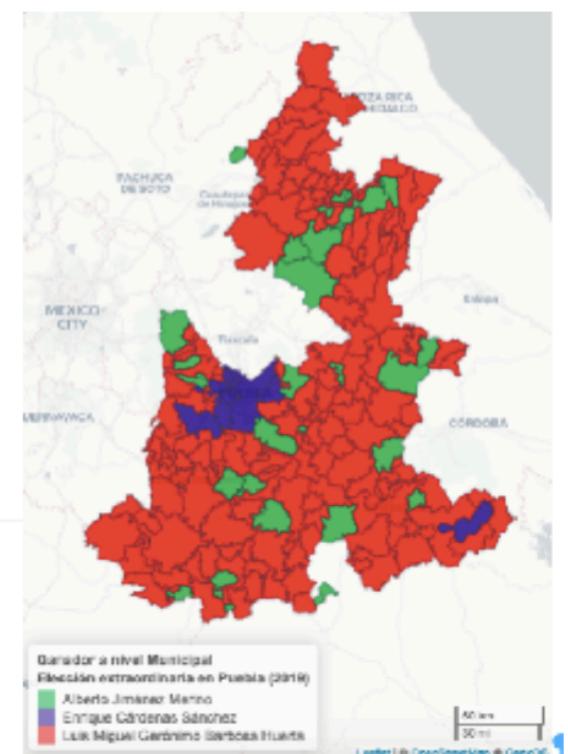


Tareas que se hacen con datos

1. Manejo y manipulación de datos en R.

`library(tidyverse)`

```
72   datos <- prep %>%
73     select(ECS, AJM, LMGBH, TOTAL_VOTOS, LISTA_NOMINAL, MUNICIPIO, DISTRITO) %>%
74     filter(!is.na(MUNICIPIO)) %>%
75     group_by(MUNICIPIO) %>%
76     summarise(ECS = sum(ECS, na.rm = TRUE),
77               AJM = sum(AJM, na.rm = TRUE),
78               LMGBH = sum(LMGBH, na.rm = TRUE),
79               Total_Votos = sum(TOTAL_VOTOS, na.rm = TRUE),
80               ListaNominal = sum(LISTA_NOMINAL, na.rm = TRUE)
81 )
```

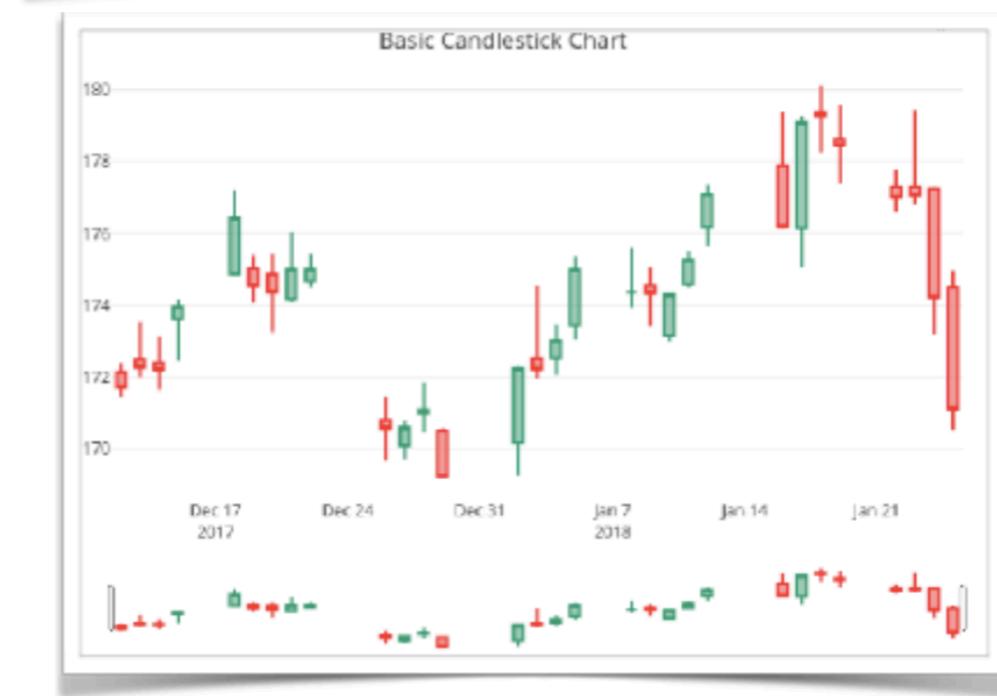
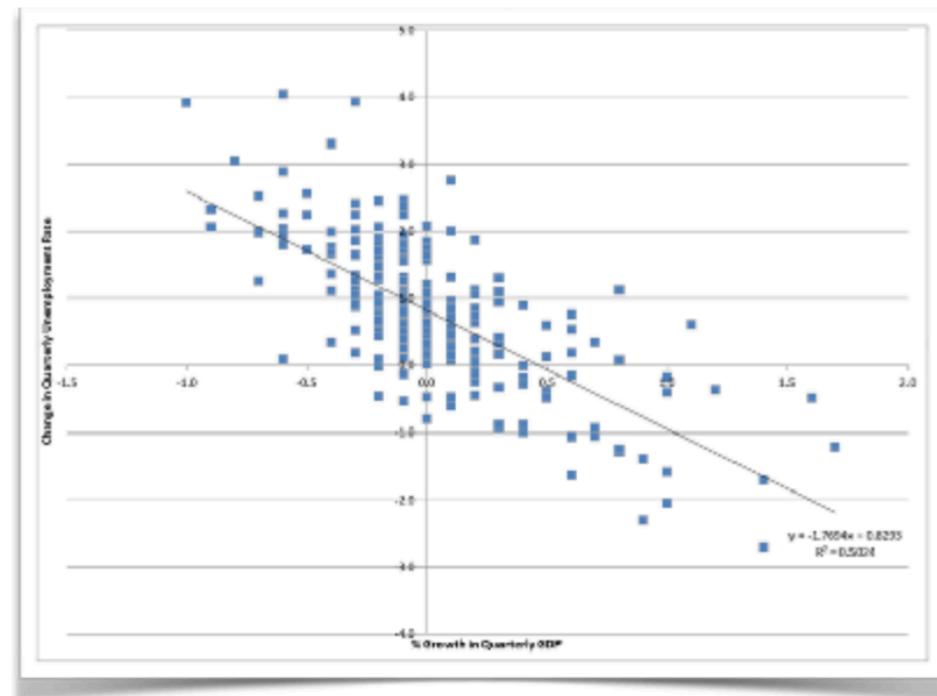
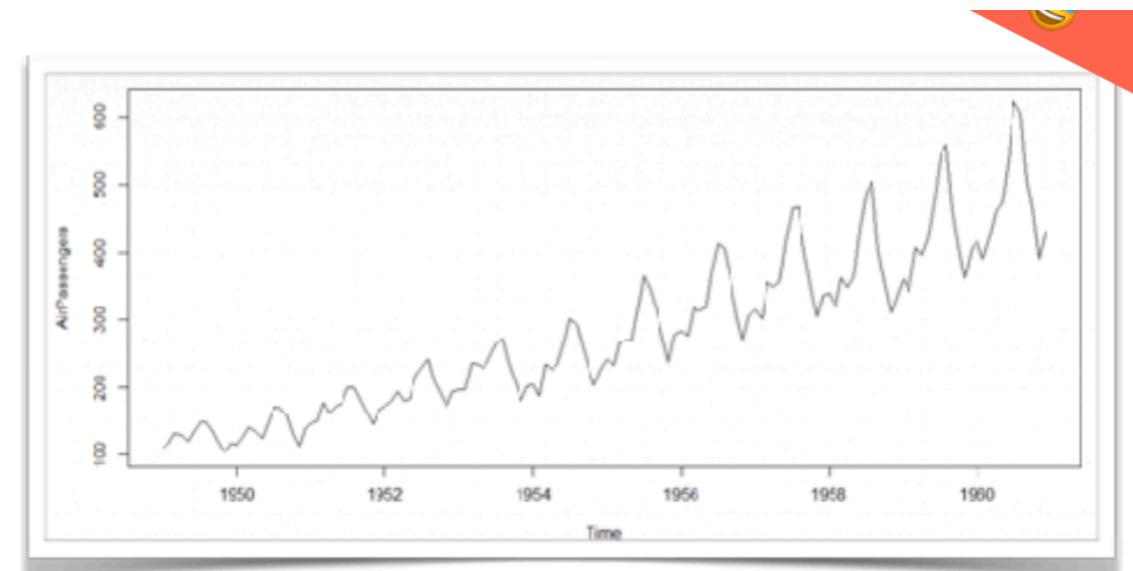


Tareas que se hacen con datos

2. Análisis estadístico y econometría.

library(base)

library(MASS)



Tareas que se hacen con datos

3. Machine Learning y Deep Learning.

`library(e1071)`

`library(tensorflow)`

`library(caret)`

`library(rpart)`

etc.

Classification



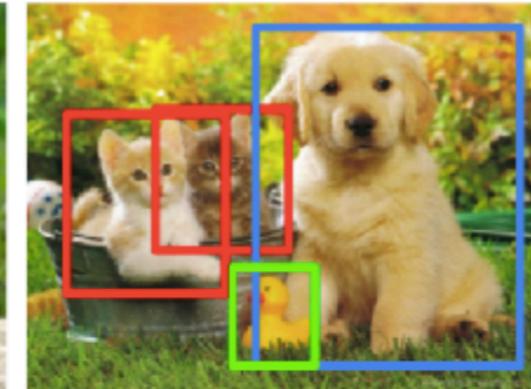
CAT

Classification + Localization



CAT

Object Detection



CAT, DOG, DUCK

Instance Segmentation



CAT, DOG, D

Single object

Multiple objects



Tareas que se hacen con datos

4. Análisis de texto.

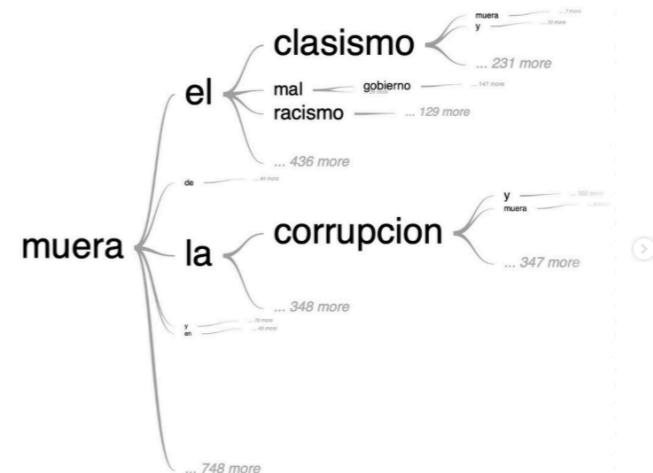
library(tm)

```
library(stringr)
```



Nube de palabras.

Solicitudes de Acceso a información realizadas en el Estado de Morelos.



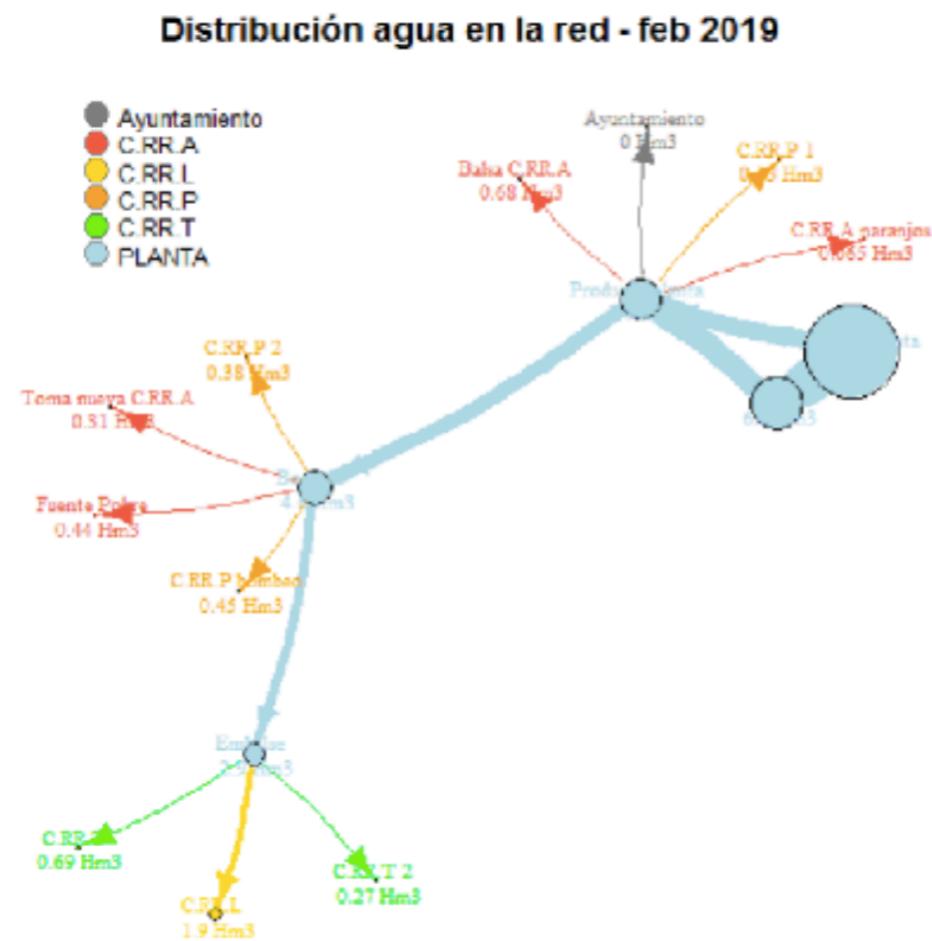
Nube de palabras.

Plan Nacional de Desarrollo. 2019.

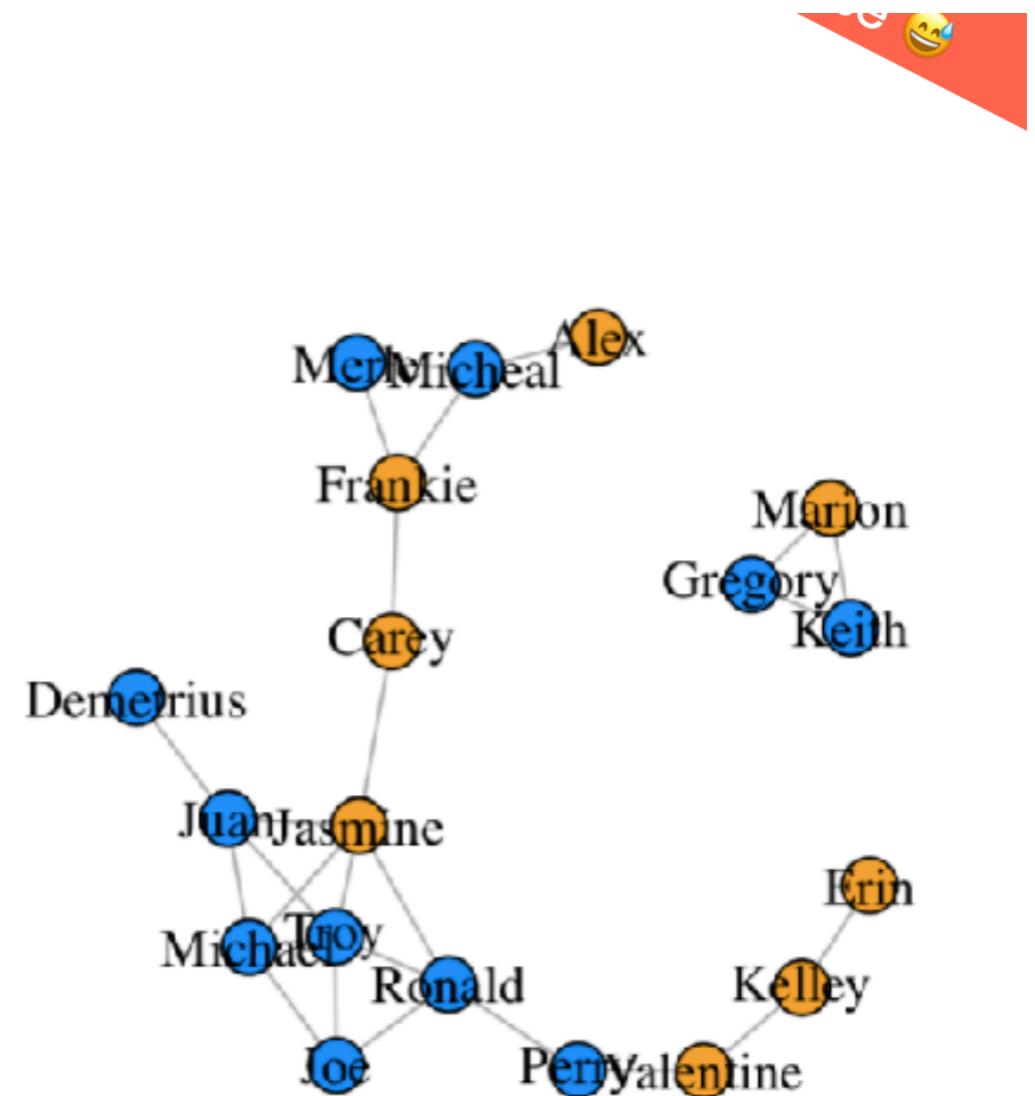
Tareas que se hacen con datos

5. Análisis de redes.

`library(igraph)`



Red de distribución de Agua



Red social de amigos en una prepa

Tareas que se hacen con datos



6. Visualización de datos.

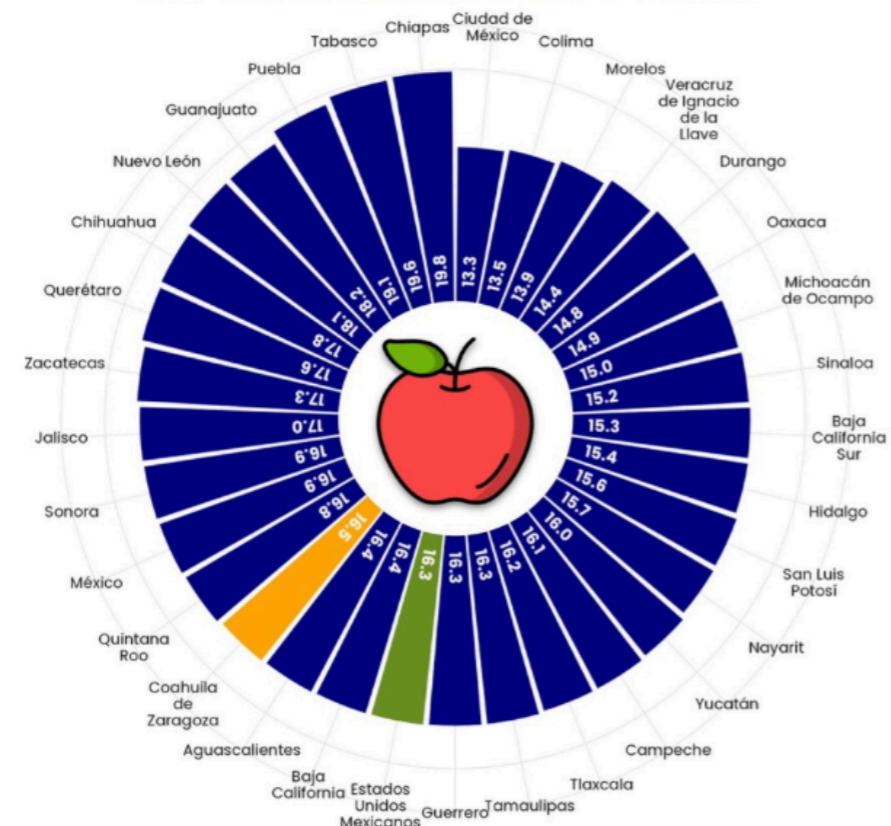
library(ggplot2)

library(plotly)

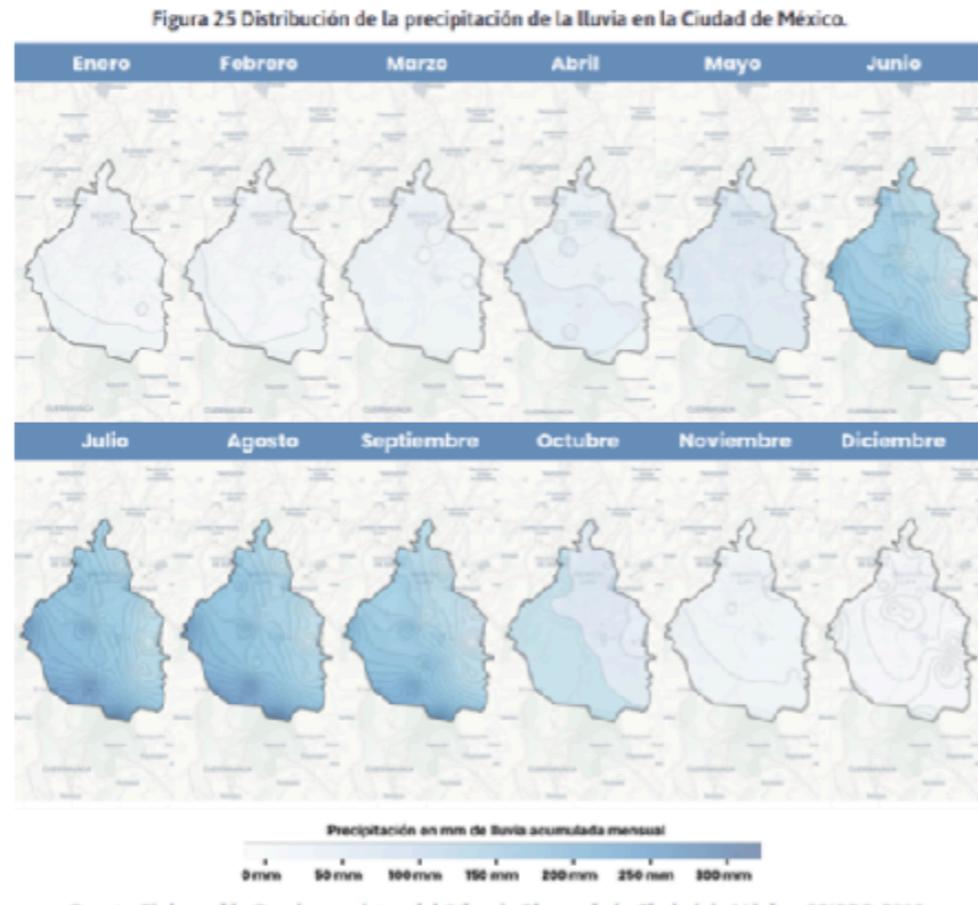
library(leaflet)

library(htmlwidgets)

Razón de **alumnos por profesor** en las entidades de México



Fuente: Elaboración propia con datos del INEGI. Ciclo 2021-2022.
@JuvenalCamposF - IG: juvenalcampos.dataviz



Fuente: Elaboración Propia con datos del Atlas de Riesgo de la Ciudad de México, SGIRPC, 2019.

**Mapas de la distribución
de la lluvia en la CDMX, en el espacio y tiempo.**

Tareas que se hacen con datos

7. Recolección automática de información (Web Scrapping, Data Crawling).

library(rvest)

library(xml)

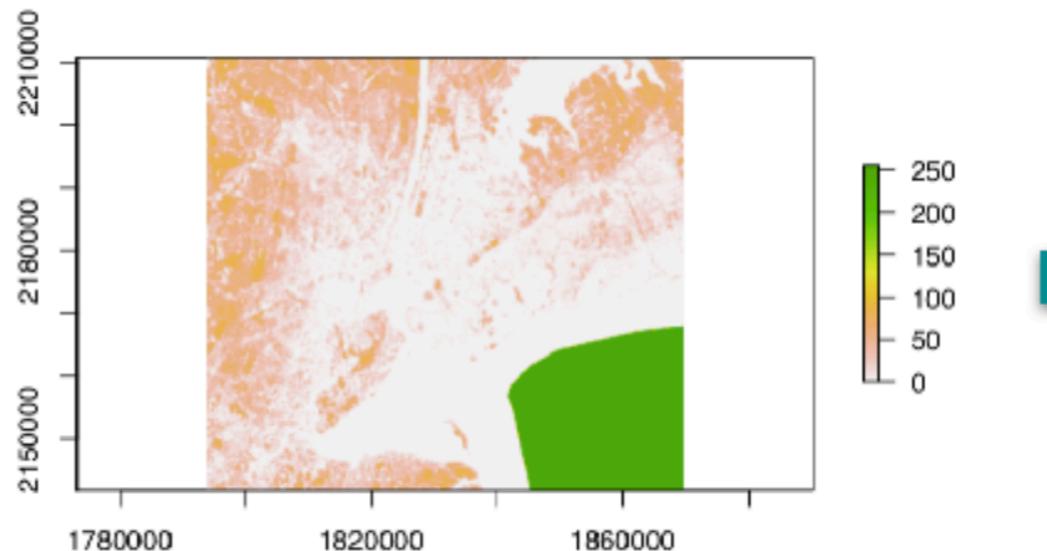
The screenshot shows the Mercado Libre website interface for used, semi-new, and new cars. At the top, there's a search bar and navigation links for categories like Ofertas, Cupones, Supermercado, Moda, etc. A banner for 'SEMI NUEVOS FARRERA' is prominently displayed. Below it, a featured listing for a blue Honda Civic 2022 is shown with a price of \$445,000. The main content area displays several car listings in a grid format, each with a thumbnail image, the vehicle's name, its condition (e.g., 'AUTO VERIFICADO'), its price, and some details about its model and year. On the left side, there are filters for 'Autos usados, seminuevos y nuevos', 'Autos y Camionetas' (with 68,474 results), and other categories like 'Tiendas oficiales' and 'Ubicación'. The bottom section shows more car models like a Mercedes-Benz Clase C and a Ford Mustang.

Tareas que se hacen con datos

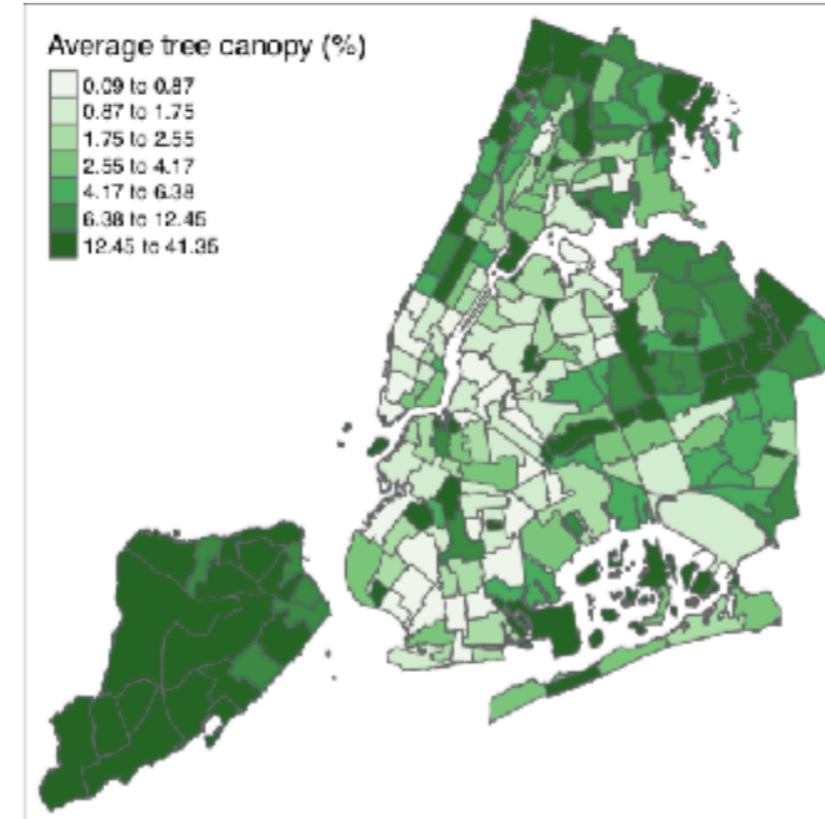
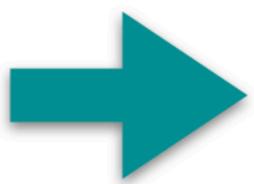
8. Análisis Geoespacial.

`library(sf)`

Abrir información geográfica, modificarla y visualizarla, así como realizar análisis a partir de esta.



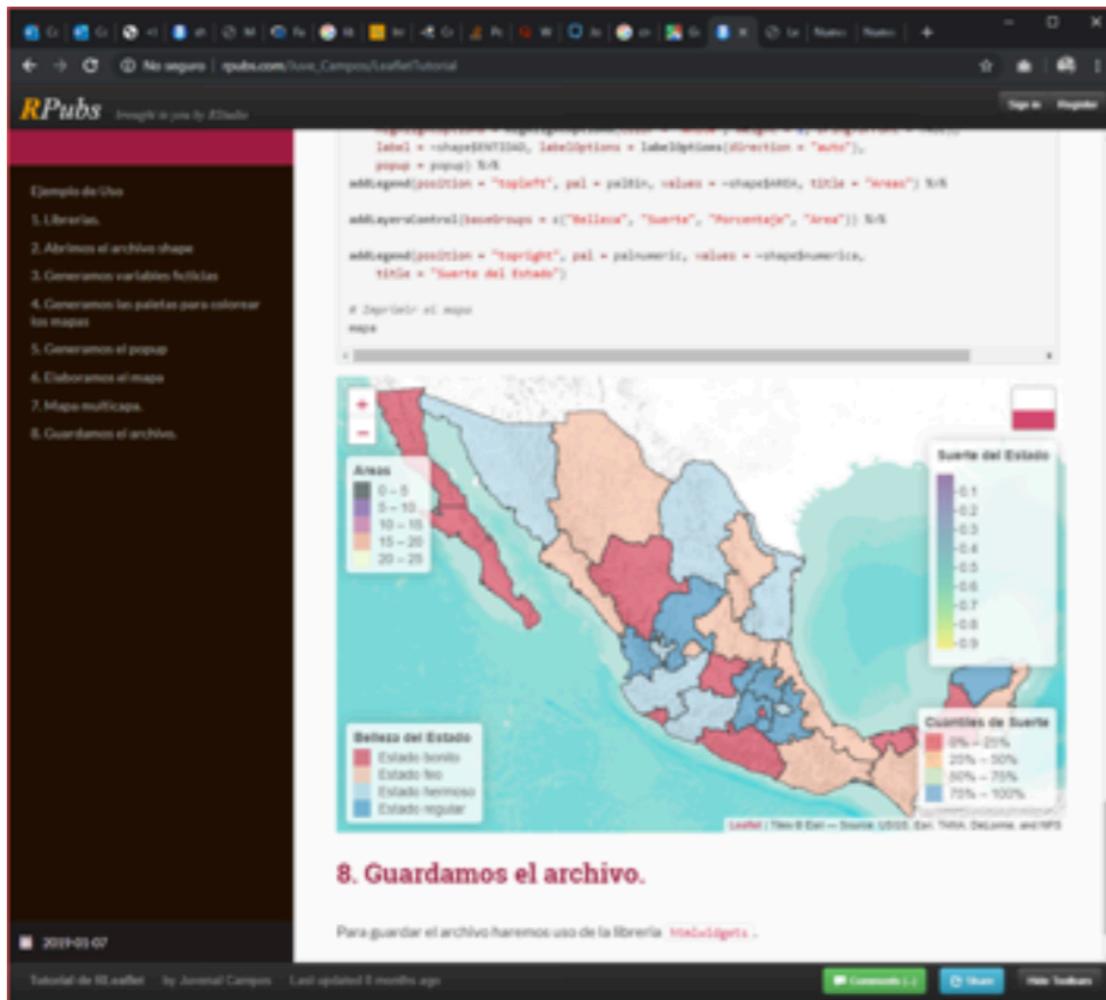
**Datos Crudos
(Raw Data)**



Datos Procesados que permiten llegar a conclusiones

Tareas que se hacen con datos

11. Páginas web y reportes (*.pdf, *.doc, diapositivas, tableros estáticos). library(markdown)



8. Guardamos el archivo.

Para guardar el archivo haremos uso de la librería [readrigner](#).

Unit 2 of the course
Juventud Campos - Based on DevelopML Classification Course
17/1/2018

Vector and raster coordinate systems

In order to perform any kind of analysis with more than one layer, you have to establish the same coordinate reference system (CRS), and the first step is determining what coordinate reference system your data has. To do this you can make use of the `sf` function `st_crs()` and the `raster` function `crs()`.

When the geodata has been read in with `sf` or `raster` has a CRS defined both sf and raster will recognize and retain it. When the CRS is not defined you will need to define it yourself using either the EPSG number or the projection.

```
#Read data
# Load the packages
library(sf)
# Linking to GDAL 3.6.1, Gdal 2.1.3, proj.4 4.9.3
library(raster)

# Loading required package: sp
root <- "/Users/juvicam/Desktop/CienciaCorta/Geospatial/Analisis_sf/R/shapefiles"
trees_sf

# Read in the tree shapefile
trees_sf <- st_read(paste0(root, "/trees/trees.shp"))

# Reading layer 'trees' from data source '/Users/juvicam/Desktop/CienciaCorta/Geospatial/Analisis_sf/R'
# with 6527 features and 7 fields
# geometry type: POINT
# dimension: XY
# bbox: -16.2500 30.4093 13.7500 30.9165
# epsg (SRID): 4326
# projection: +proj=longlat +ellps=WGS84 +no_defs
# Read in the neighborhood shapefile
neighborhoods <- st_read(paste0(root, "/neighborhoods/neighborhoods.shp"))

# Reading layer 'neighborhoods' from data source '/Users/juvicam/Desktop/CienciaCorta/Geospatial/Analisis_sf/R'
# with 128 features and 8 fields
# geometry type: POLYGON
# dimension: XY
# bbox: -16.2500 30.4093 13.7500 30.9165
# epsg (SRID): 4326
# projection: +proj=longlat +ellps=WGS84 +no_defs
# Read in the tree canopy raster
canopy_sf <- raster(paste0(root, "/canopy.tif"))
```

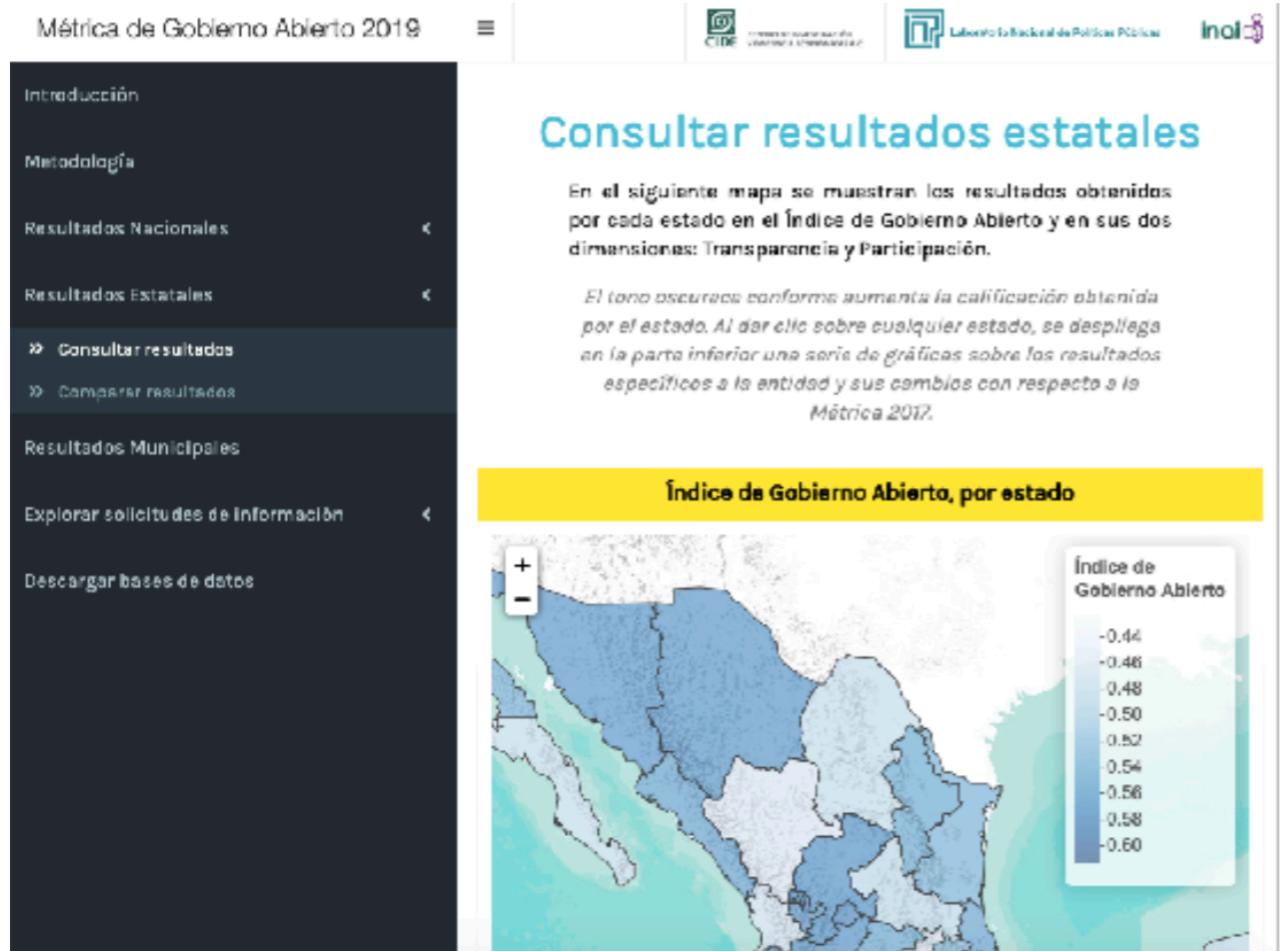
http://rpubs.com/Juve_Campos/LeafletTutorial

IATEX pdf

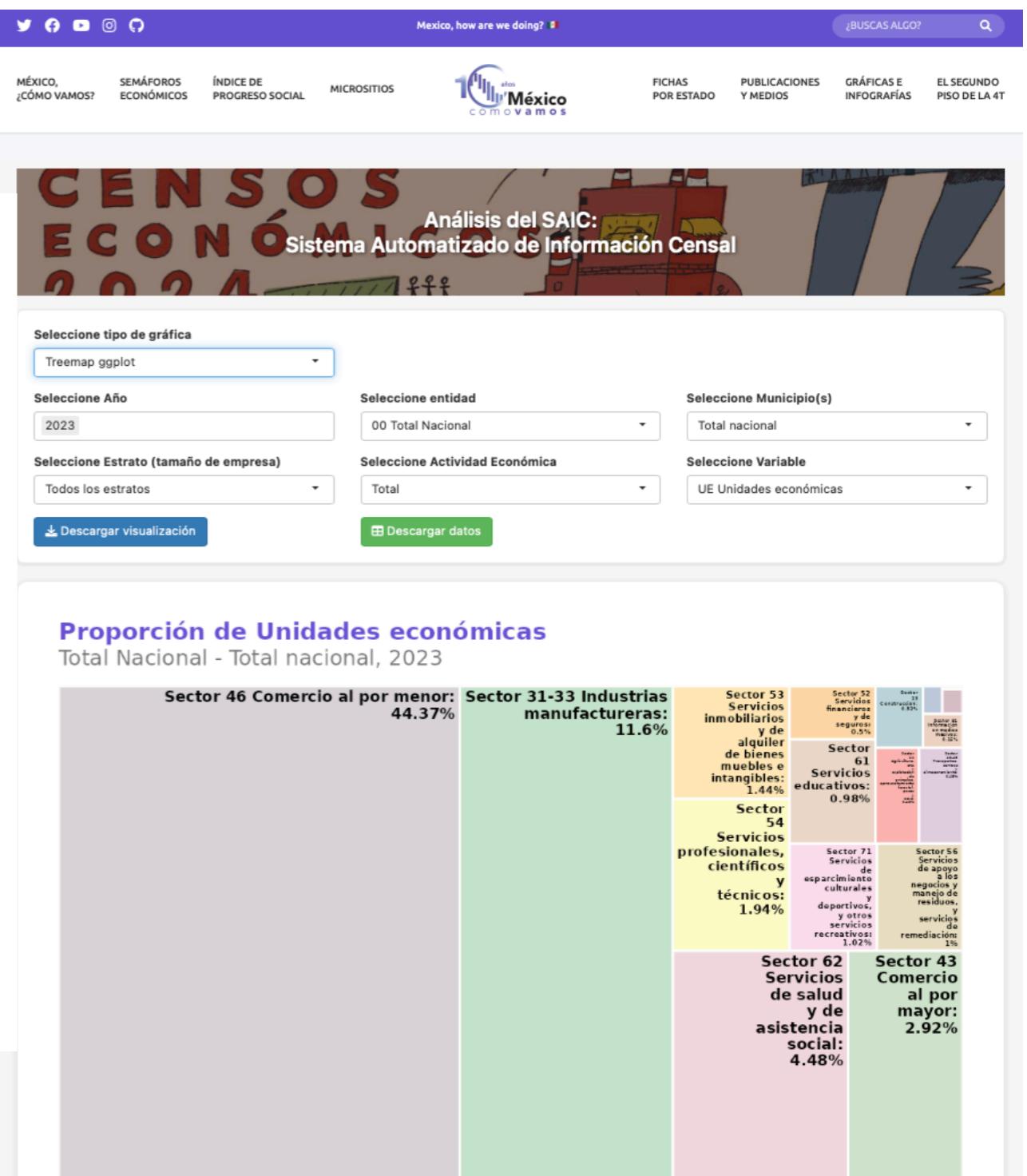
Tareas que se hacen con datos

12. Web Apps.

library(shiny)



Página web de resultados de la métrica de gobierno abierto, 2019.



Tarea

1. En equipo, elabore, usando LLMs y a partir de lo visto en clase, una guía de estudio para todo el contenido del temario.