



# Proyecto Corto 2

23/11/2020

---

Ana María Guevara Roselló. 2018102514

Luis Alejandro Masís Pérez. 2017239616

Instituto Tecnológico de Costa Rica. Ingeniería en Computación

Visualización de la Información

Lilliana Sancho Chavarría

## Índice

<b>Índice</b>	<b>1</b>
<b>Introducción</b>	<b>1</b>
<b>Objetivo general</b>	<b>2</b>
<b>Especificaciones</b>	<b>2</b>
<b>Aspectos técnicos</b>	<b>2</b>
Herramienta de desarrollo	2
Algoritmos	3
<b>Programa</b>	<b>3</b>
Bienvenida y configuración	3
Visualización y gráfico	6
<b>Aspectos Visuales y diseño del proyecto</b>	<b>7</b>
Escalas de Colores Utilizadas	8
<b>Análisis de Resultados</b>	<b>8</b>
¿Qué aprendimos?	8
Partes Completadas	9
Cuadro de Tareas	11
<b>Bibliografía</b>	<b>12</b>

## Introducción

Un gráfico de coordenadas paralelas se utiliza para trazar datos varios, en donde su intención principal es mostrar la relación entre sí y sus intersecciones, a su vez comparar estos datos. Este tipo de visualización se utiliza para trazar datos numéricos de valores variados. Además, son ideales para comparar muchas variables y ver sus relaciones.

En un gráfico de coordenadas paralelas cada una de las variables se relaciona con los ejes definidos para la visualización. La escala de estos ejes puede ser variada, ya que es de acuerdo al valor de los datos de cada eje. Sin embargo todas las escalas pueden normalizarse definiendo un rango de valores igual para cada uno de los ejes.

La desventaja de este tipo de visualización es que al tener muchos datos la información se ve aglomerada e ilegible. Aunque pueden aplicarse técnicas para corregir esa ilegibilidad de los datos.

## Objetivo general

Desarrollar una herramienta configurable e interactiva para la creación de gráficos de coordenadas paralelas.

## Especificaciones

Este proyecto consiste en desarrollar una herramienta configurable e interactiva para visualizar información mediante gráficos de coordenadas paralelas. La herramienta debe tener las siguientes características:

- A. Leer los datos de un archivo y visualizar el gráfico.
- B. Una función que permita al usuario escoger la forma en que desea que se visualicen los datos, ya sea si mantiene los datos en la escala original o si normaliza los datos.
- C. Modificar el orden en que están dispuestos los ejes. Idealmente que el usuario interactúe con la herramienta y que mueva los ejes a otras posiciones mediante algún mecanismo de interacción que usted defina.
- D. Implementar la técnica de cepillado.
- E. Configurar los colores de:
  - las variables (es decir, que el usuario pueda escoger los colores de una paleta de colores y/o individualmente)
  - Configurar los colores de los ejes.
- F. Configurar el grosor de
  - las líneas de datos

- las líneas de los ejes
- G. Seleccionar al menos dos conjuntos de datos para probar la herramienta.
  -

## Aspectos técnicos

### Herramienta de desarrollo

Como lenguaje de desarrollo se utilizó Python y para la graficación se utilizó la librería Plotly, específicamente el paquete “graph\_objects”

Esta librería permite una visualización de gráficos interactiva. Utilizando como base para generar la visualización del proyecto los ejemplos de esta librería, los cuales se pueden encontrar aquí:

[Parallel Coordinates Plot | Python | Plotly](#)

### Algoritmos

Para la extracción de datos del archivo Excel, se utilizó un doble ciclo en el que cada dato es agregado a un diccionario, en donde el nombre de la columna es su llave de búsqueda y esta contiene todos los elementos en la columna, de forma:

```
“ {'Model': ['chevrolet chevelle malibu', 'buick skylark 320',...], 'MPG': [18.0, 15.0, 18.0...],etc} ”
```

Esto debido a que el conjunto debe ser cargado a la visualización por columna. Este proceso se puede encontrar en el archivo “Excel.py”. Una vez extraída la información, se pasa al proceso de visualización.

En “Visualizacion.py” encontraremos un doble ciclo que ordenará los datos generados anteriormente para visualizar, generando a su vez un diccionario para cada columna, donde se especifica el contenido de sus columnas y el nombre de ellas.

En caso de que los datos sean normalizados es necesario llevar a cabo otros algoritmos para sacar los valores mínimos y máximos de todos los datos, en caso contrario, se lleva a cabo el ciclo con normalidad. A su vez, también se hace la distinción en caso que los valores no sean numéricos, es decir, sean palabras, si es así entonces no se aplicará la normalización.

Una vez extraídos los valores de las columnas, estos se envían a generar la visualización, utilizando Parcoords, en donde se terminan de configurar las especificaciones visuales y de datos.

Pese a ser una herramienta flexible, no permite una completa personalización de los gráficos en todas sus funciones, por ejemplo, el ancho de las variables y columnas, que inicialmente se encontraban dentro del proyecto.

Para llevar a cabo la visualización es necesario tener instalada la librería Plotly y Python 3. Estos pueden ser descargados siguiendo estos pasos:

[Plotly With Python : 7 Steps - Instructables](#)

## Programa

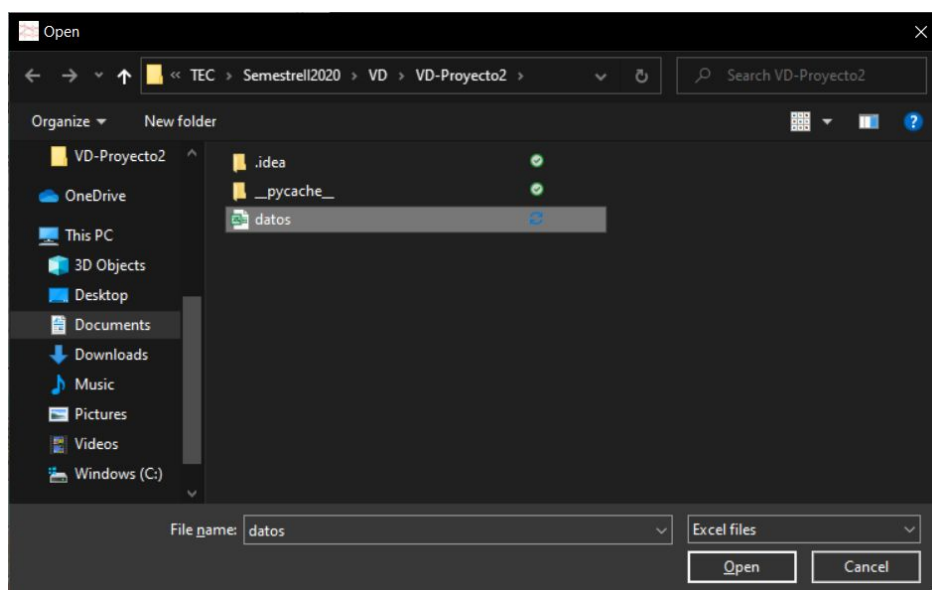
### Bienvenida y configuración



Para ejecutar el programa únicamente es necesario ejecutar el archivo "main.py".

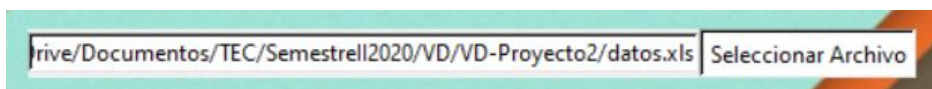
Al iniciar el programa una pantalla de bienvenida aparecerá. En esta se podrán seleccionar las diferentes configuraciones de la visualización de coordenadas paralelas.

La primer configuración es la única que no está predefinida, es necesario presionar el botón "Seleccionar Archivo" y seleccionar un archivo de formato .xls



El buscador de archivos que se desplegará una vez se seleccione el botón, a su vez, solo mostrará carpetas y aquellos archivos que cumplan con el formato establecido.

Una vez seleccionado el archivo, se mostrará su dirección en el cuadro de texto contiguo al botón.



La siguiente configuración realizable será la de escala de colores de las variables. Esta estará predefinida como magma.

Esta configuración se divide en 2 posibilidades:

La primera es su edición utilizando el conjunto de opciones de escalas, que contienen las escalas de llamadas "Magma", "Solar", "Thermal", "Haline" y "Personalizada".

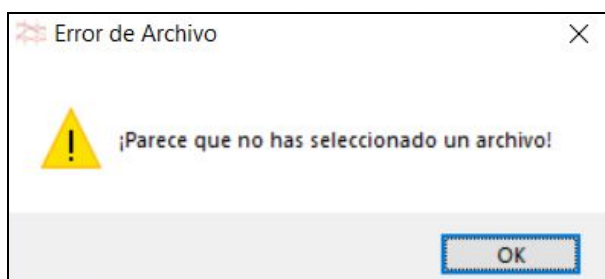
La segunda posibilidad es seleccionar el color que desea editar dentro de la escala y modificarlo. Una vez seleccionado el color aparecerá en pantalla un seleccionador de colores y automáticamente su selección de escala de colores se cambiará a "Personalizada".

La tercera configuración consiste en el color de las columnas del gráfico, que utiliza la misma dinámica anterior para seleccionar un color.



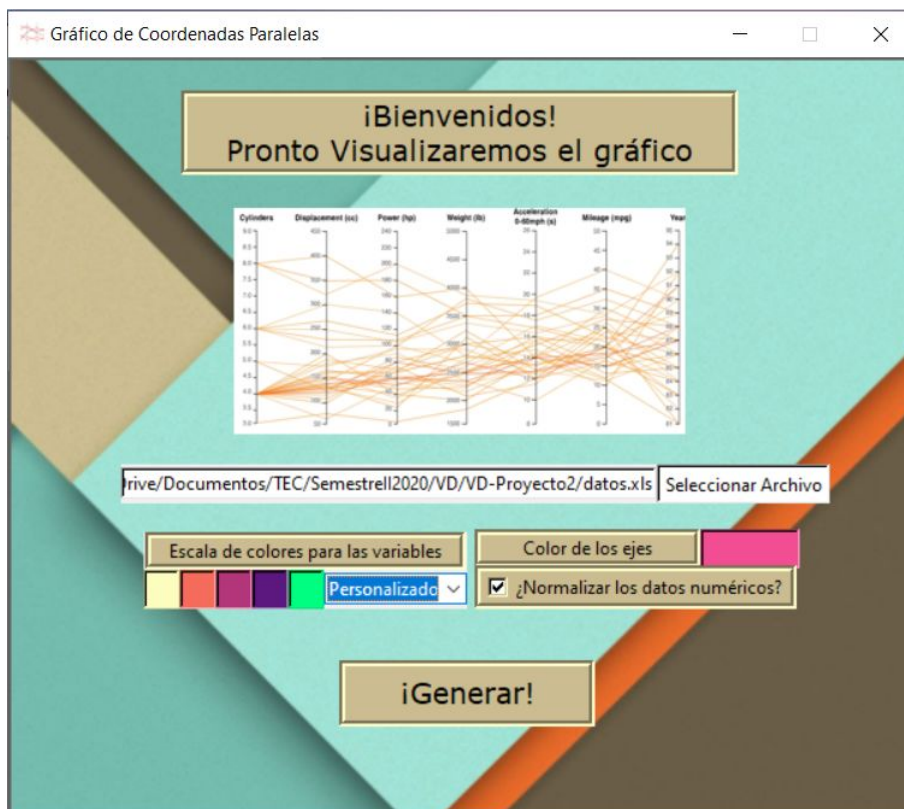
La cuarta configuración consiste en una casilla de selección. Cuando el usuario desee ver los datos normalizados deberá seleccionar esta casilla, en caso contrario, deberá dejar la casilla en blanco.

La visualización podrá verse al seleccionar el botón “Generar”. Esto sólo funcionará si se ha seleccionado un archivo válido. En caso contrario, mostrará una ventana de alerta y no permitirá la visualización. Cualquiera de los otros valores de modificación pueden ser dados por sentado y la visualización se generará sin problemas.



## Visualización y gráfico

Al generar la visualización exitosamente se abre una ventana en el buscador predeterminado, que mostrará lo generado por las configuraciones.



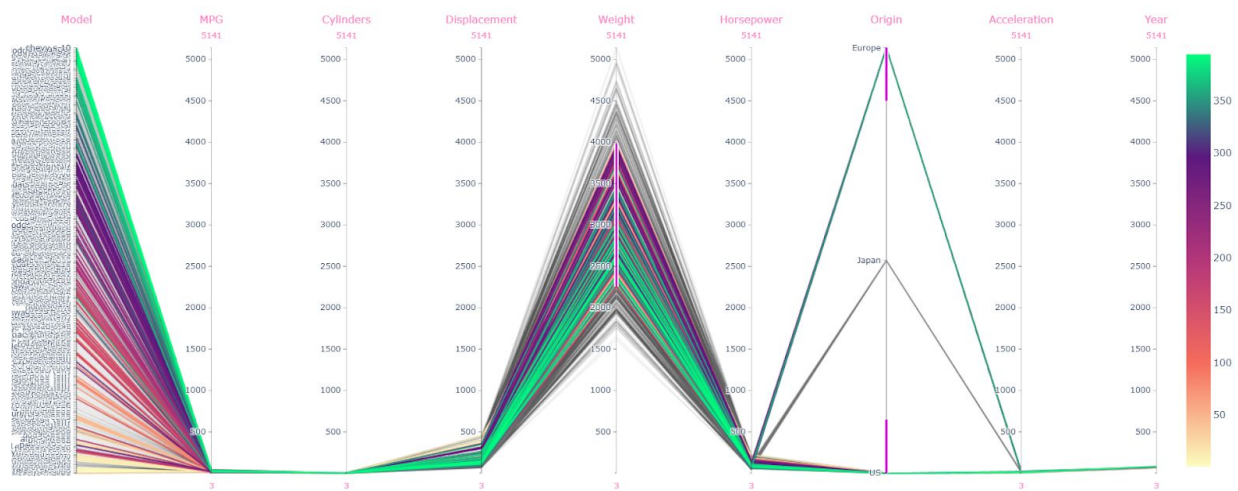


Al mostrarse la visualización podremos observar los datos de la siguiente forma, según el ejemplo anterior:

Como podemos ver, el gráfico obtenido tiene la posibilidad de mover sus columnas a voluntad del usuario, además de filtrar los datos utilizando barras de selección en las mismas.

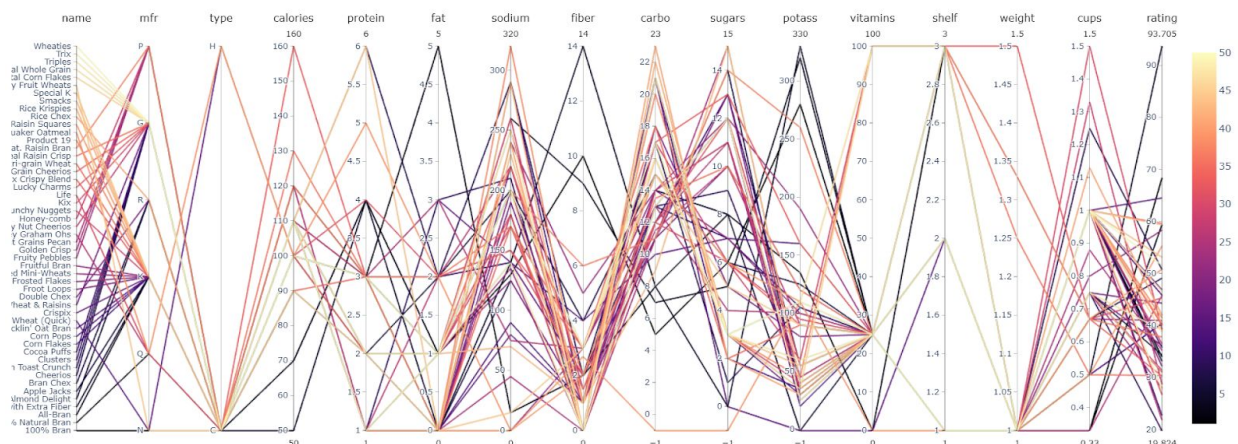
Para mover las columnas, solo se debe posicionar el mouse sobre el nombre de la columna y arrastrarla a su posición deseada. Mientras que para filtrar los datos, es necesario seleccionar o arrastrar el mouse por sobre la columna y el rango que se desea filtrar. Además, se pueden hacer múltiples filtrados.

Para borrar uno de los filtros, se puede hacer click sobre el rango de este filtro.



El rango de visualización óptimo de datos es de 1 a 85 datos en este ejemplo.

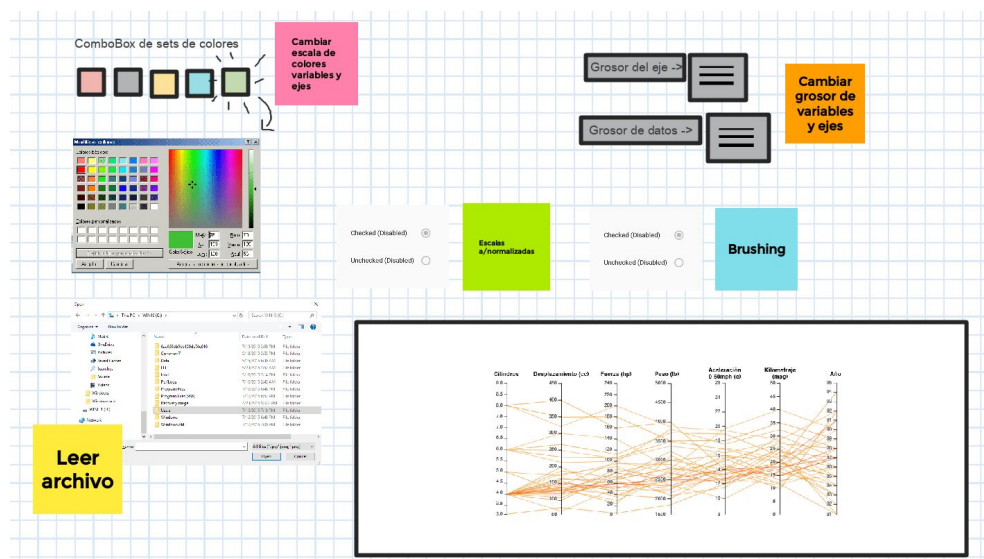
Visualización generada con el conjunto de datos "cereal.xls" utilizando la escala de colores "Magma":





## Aspectos Visuales y diseño del proyecto

Como diseño inicial se hizo un mapa de las funcionalidades necesarias para el proyecto:



Se eliminó la selección de brushing debido a la visualización proporcionada por plotly.

## Escalas de Colores Utilizadas



**Magma:** que consiste en un conjunto de colores seleccionados usando el método de triada(en quintos) y pasa de un tono oscuro a uno claro.



**Solar:** Consiste en un conjunto de colores seleccionados por adyacencia, de nuevo, utilizando tonalidades oscuras y luego claras.



**Thermal:** Un conjunto de colores generados de forma complementaria. Que abarca un esquema de azul oscuro a un amarillo cadmio.

**Haline:** Un conjunto de colores análogos entre los colores primarios azul y amarillo.



## Análisis de Resultados

### ¿Qué aprendimos?

Como en toda visualización, los detalles son importantes, por lo que aspectos como la selección de colores, tamaño de letra, distribución del espacio en pantalla y demás, necesitan tomarse muy en cuenta. En la parte visual como tal aprendimos a mejorar el aspecto tomando en cuenta lo aprendido en el curso, por ejemplo, la distribución de la pantalla inicial de la interfaz, los colores utilizados en esta y el tipo de letra. Además, la forma en que se muestra el gráfico, sus colores y su distribución fueron previamente analizados, por lo que tomar en cuenta detalles que al final eran relevantes fue algo que tuvimos que aprender.

Como parte de los requerimientos del proyecto debía cargarse un archivo con los datos que iban a visualizarse previamente, por lo que se investigó la forma de obtenerlos y manejarlos convenientemente para uso previo. También se investigó el uso de la herramienta que permite graficar los datos obtenidos.

En general el proyecto deja un alto nivel de aprendizaje, principalmente en la evaluación en grupo de cómo hacer las cosas, cómo darle a la visualización un mejor aspecto y qué es mejor eliminar o no de esta. Por lo que aplicar los conocimientos adquiridos durante el curso aporta un aprendizaje extra, debido a que se evalúan mejor los aspectos relacionados con la parte visual.

### Partes Completadas

#### A. Leer datos de un archivo (Porcentaje de realización: 100%)

Se implementó un explorador de archivos, el cual obtiene la dirección en memoria de este archivo, luego envía esa dirección a una función que recorre este archivo generando un diccionario que tiene como llaves los nombres de los encabezados de columna y como valores de estas llaves todos los datos que tiene esa columna.

B. Permitir al usuario normalizar los datos (Porcentaje de realización: 100%)

En la pantalla inicial de la interfaz el usuario, mediante un “checkboxbutton”, puede seleccionar la opción de que los datos se normalicen para que en la visualización todos los ejes estén en el rango normalizado.

C. Configurar el color de las variables con paletas predefinidas (Porcentaje de realización: 100%)

Al iniciar el programa el usuario puede seleccionar una de las cuatro paletas de colores predefinidas para la visualización.

D. Configurar el color de las variables de forma individual (Porcentaje de realización: 100%)

Además de las cuatro paletas predefinidas, el usuario puede seleccionar la opción de “Personalizado”, en la cual pueden indicar cinco colores distintos mediante un “colorchooser”. Estos cinco colores son los que se van a tomar en cuenta para visualizar el gráfico.

E. Configurar el color de los ejes (Porcentaje de realización: 100%)

Esta opción se ejecuta similar al punto d). En la interfaz se muestra la opción de elegir un color para los ejes, el cual se obtiene mediante un “colorchooser”.

F. Configurar el grosor de las líneas en los datos y ejes (Porcentaje de realización: 0%)

Debido a que la herramienta que se empleó para la realización del proyecto no permitía cambiar estos valores, la realización de este punto no se logró en su totalidad.

G. Visualizar el gráfico (Porcentaje de realización: 100%)

La visualización del gráfico se completó en su totalidad, pues esta muestra correctamente las relaciones entre los ejes, la cantidad total de datos y sus valores.

#### H. Modificar el orden de los ejes (Porcentaje de realización: 100%)

El usuario puede seleccionar uno de los ejes y re acomodarlo en cualquier posición a derecha o izquierda. Al mover los ejes los demás datos y ejes no se ven afectados.

#### I. Implementar la técnica del cepillado (Porcentaje de realización: 75%)

Para la técnica del cepillado se pueden filtrar datos en todos los ejes, por ejemplo, se selecciona un valor en el primer eje, uno en el segundo, hasta el eje n, visualizando sólo las variables que pasan por esos puntos específicos. Pero debido a que no se pudo solucionar el hecho de que cuando la cantidad de datos de un eje es muy grande se aglomeran todos. Por esta razón no se completó la tarea en un 100%, debido a que se limita la visualización de los datos cuando estos son muchos.

### Cuadro de Tareas

Tarea	Porcentaje de realización
A	100%
B	100%
C	100%
D	100%
E	100%
F	0%
G	100%
H	100%
I	75%

## Bibliografía

Hektor Profe: Interfaces gráficas con Tkinter. (4 de octubre de 2018). Recuperado de <https://docs.hektorprofe.net/python/interfaces-graficas-con-tkinter/>

Python: 24.2 tkinter.ttk - Tk themed widgets. (12 de octubre de 2014). Recuperado de <https://docs.python.org/3.2/library/tkinter.ttk.html#module-tkinter.ttk>

Weitz, D. (27 de julio de 2020). Parallel Coordinates Plots. towards data science. Recuperado de

<https://towardsdatascience.com/parallel-coordinates-plots-6fcfa066dcb3>

IEEE. (15 de julio de 2009). Visual Perception of Parallel Coordinate Visualizations. IEEE Xplore. Recuperado de <https://ieeexplore.ieee.org/document/5190777>

plotly.graph\_objects.Parcoords. (s.f). Recuperado de [https://plotly.github.io/plotly.py-docs/generated/plotly.graph\\_objects.Parcoords.html](https://plotly.github.io/plotly.py-docs/generated/plotly.graph_objects.Parcoords.html)

Parallel Coordinates Plot in Python. (s.f). Recuperado de <https://plotly.com/python/parallel-coordinates-plot/>

Python Figure Reference: parcoords Traces. (s.f). Recuperado de <https://plotly.com/python/reference/parcoords/>