

Visualización de Datos Avanzada

Análisis de Datos con Python

Eduardo Selim Martínez Mayorga



- Aprender a modificar los estilos predeterminados de nuestras gráficas
- Conocer y aprender a interpretar las siguientes gráficas:
 - Treemaps
 - Scatterplots por categorías
 - Scatterplots con variables condicionantes
 - Binnings Hexagonales
 - Mapas cloropléticos
 - Gráficas de barras apiladas



- Poblaciones y muestras
- Sesgos
- Distribuciones muestrales de estadísticas, bootstrap
- Error estándar e intervalos de confianza
- Técnicas de evaluación de modelos
- Entrenamiento y prueba
- Validación cruzada

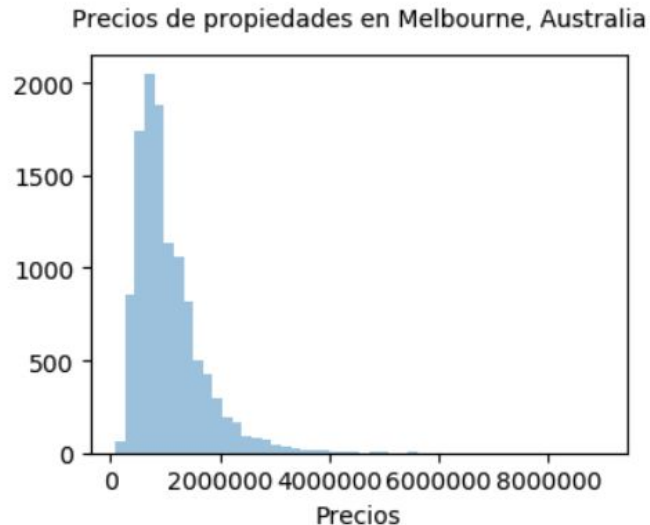
Antes de empezar



¡No olvides hacer pull del repo!

El material de la sesión se encuentra ahí.

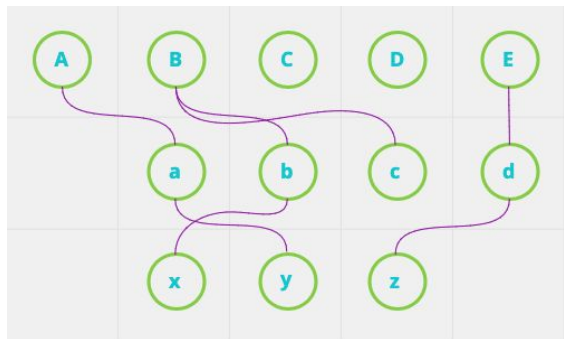
```
git pull origin master
```



Tanto seaborn como matplotlib traen varios estilos precargados que podemos aplicar a nuestras gráficas para cambiar su apariencia.

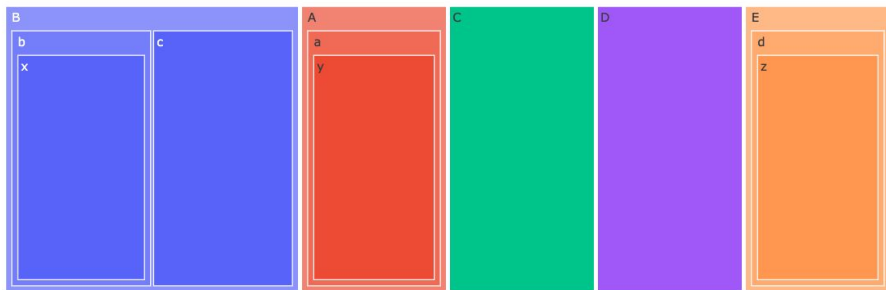
Veamos algunos de estos estilos.

[Vayamos al Repo](#)



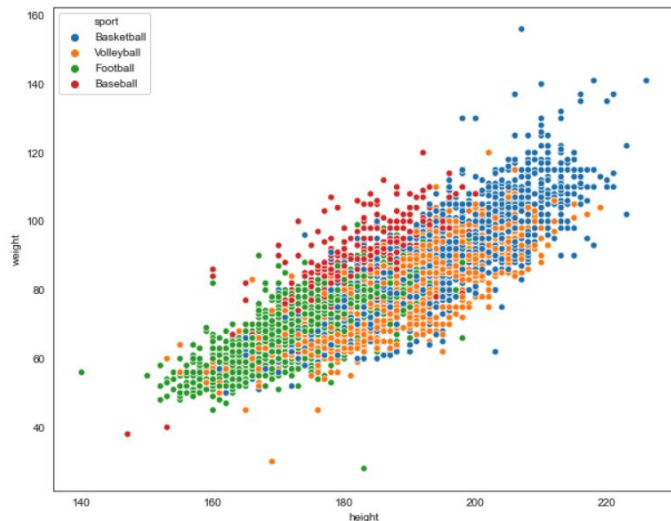
Los treemaps son gráficas que sirven para visualizar datos jerárquicos usando figuras (normalmente rectángulos) anidadas.

Si pensamos en cada nivel como una columna categórica en un DataFrame, podemos construir un treemap usando rectángulos



Vayamos al Repo

Dispersión por categorías

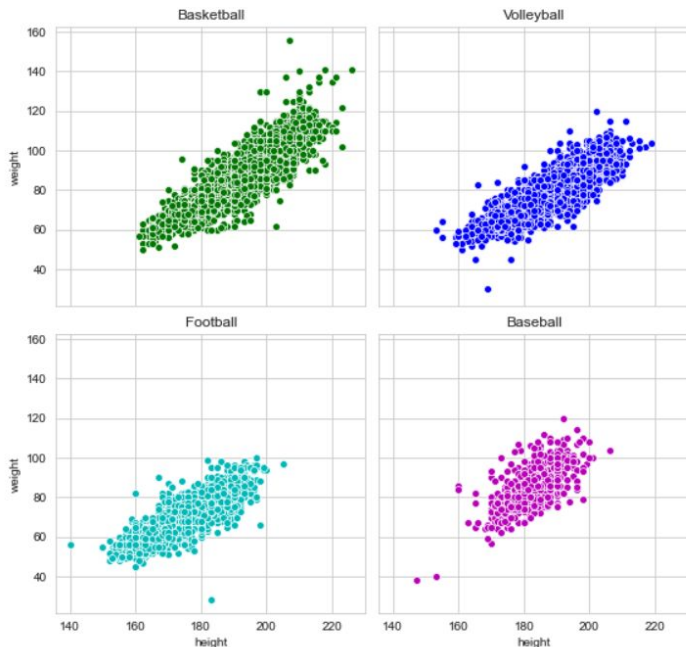


- Necesitamos tener al menos dos variables numéricas
- Coloreamos los puntos usando una variable categórica

Vayamos al Repo

Dispersión por Variables Condicionantes

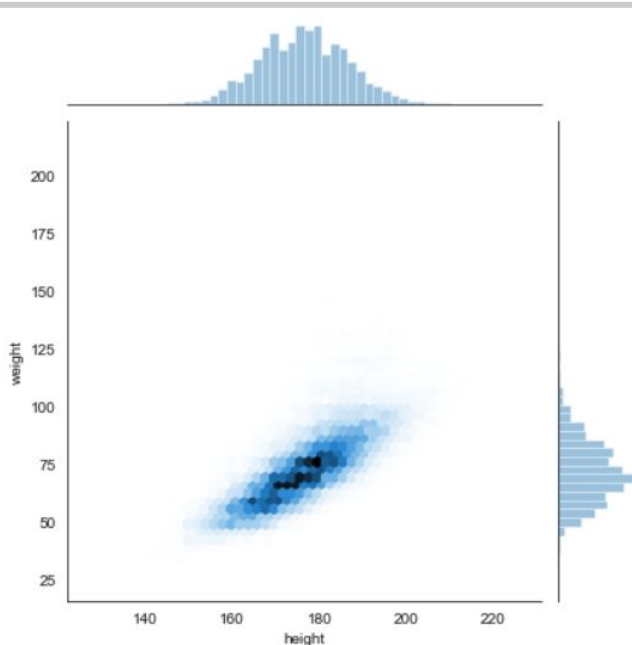
Relationship between height and weight in 4 sports



- La variable (o variables) condicionante es la variable categórica que quieres comparar.
- Para hacer este tipo de gráfica primero tenemos que separar nuestro dataset en dos o más partes.
- Después realizamos múltiples gráficas, una por cada valor de la variable categórica, los colocamos en distintas gráficas y las comparamos.

[Vayamos al Repo](#)

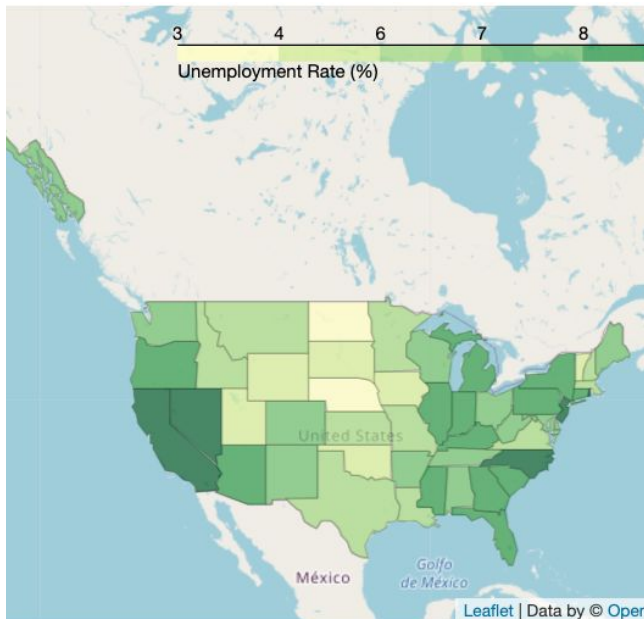
Binning Hexagonal



Vayamos al Repo

- Divide el plano cartesiano en hexágonos.
- Coloca los puntos del scatterplot sobre el plano.
- Cuenta cuántos puntos caen en cada hexágono.
- Le asigna a cada hexágono un color dentro de un espectro. Entre más oscuro sea el color, más puntos había sobre de ese hexágono
- Agrega el histograma de cada variable.

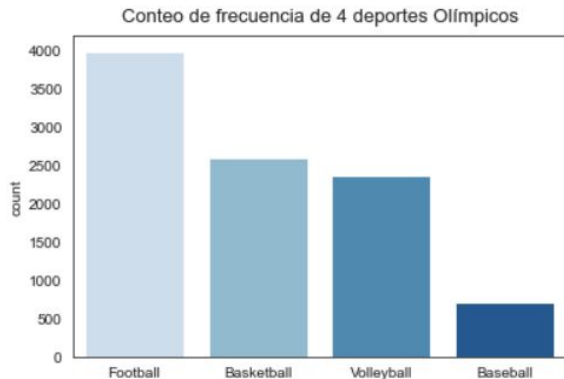
Mapas Cloropléticos



- Primero dividimos nuestro mapa en regiones.
- Luego coloreamos cada región de acuerdo a un espectro de color que representa la segmentación de una variable estadística.
- Para ubicar las regiones suelen usarse archivos GEOJson

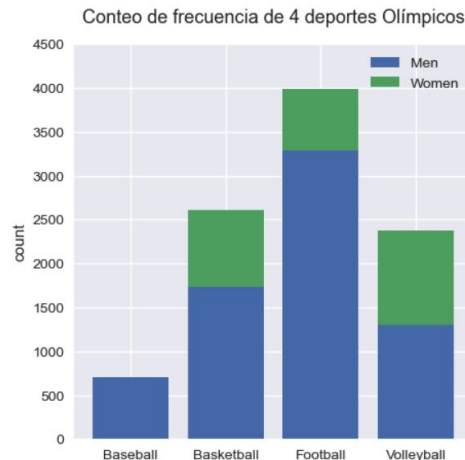
Vayamos al Repo

Gráficas de barras apiladas



Vayamos al Repo

- Nos permiten mostrar un porcentaje con respecto a una variable categórica.
- Al igual que en la dispersión por categorías, necesitamos variables numéricas.





NO OLVIDES REVISAR TU
POSTWORK Y TU PREWORK



Preguntas

