

# Visualización de la Información



Finanzas 2019  
Pablo Alejandro Costesich  
Legajo #50109

# Agenda

- Objetivo
- Data extraction & wrangling
- Visualizaciones
- Dificultades encontradas
- Conclusiones

---

# Intro

# Objetivo

Utilizando la información de mi cuenta personal, entender el flujo de fondos

Se obtuvo data de:

- Cuenta Corriente (gastos en débito)
- Cotización del dólar
- ...y frecuencia cardíaca basal. :)

---

# Etapas Planteadas

1. Adquisición de datos
  - a. Mediante la data de Home Banking
  - b. HealthKit de Apple (Apple Watch)
2. Data Wrangling
  - a. Procesar los datos de distintos formatos para obtener CSVs
    - i. El HealthKit Export viene en formato XML
    - ii. La data del Banco Galicia viene en XLSX
  - b. Anonimizar los datos
3. Presentar y analizar la información
  - a. Diagrama de distribución
  - b. Evolución temporal
  - c. Sub-segmentación

**Data extraction**

**Data wrangling**

# Data Extraction

## Banco Galicia

La data del banco está limitada a las siguientes condiciones:

- Sólo gastos con débito. Eso incluye:
  - el pago (total) de las tarjetas
  - débitos automáticos de servicios
- Sólo los últimos 3 meses tienen toda la información.
- Una única cuenta corriente

## HealthKit

Apple provee una forma de exportar toda la data de HealthKit mediante un ZIP, pero...

- Contiene los datos crudos de muchas métricas...
- ...que a menudo son redundantes.
- Porque también contiene los agregados.

# Data Extraction - Banco Galicia

Hice un script que convierte el XLSX exportado a CSV.

El script tiene algunas funciones de data wrangling:

- Anonimiza los montos: convierte a valores en porcentajes según el máximo histórico del balance de la cuenta (si el máximo fue de \$100000, un gasto de \$5000 aparece como 5%).
- Agrupa en categorías: para hacer mejores comparativas, agrega los datos según se gastó en servicios, tarjeta, comida, supermercado, transporte u ocio.



# Data Extraction - Banco Galicia

**Hice un script que convierte el XLSX exportado a CSV.**

El script está hecho en NodeJS, y la salida es un CSV por consola. El único parámetro de entrada es el nombre del archivo Excel con el estado de cuenta.

# Data Extraction - HealthKit

La extracción de HealthKit fue más rudimentaria. En vez de utilizar un script de Node, opté por hacer un script de Bash y manualmente procesar el texto.

La métrica interesante es ***RestingHeartRate***. Es la frecuencia cardíaca basal, que a menudo se utiliza para indicar estrés.

La extracción se realizó con la siguiente línea:

```
grep RestingHeartRate export.xml | cut -d '"' -f 10,16 | sed 's/"/, /g' >hr.csv
```

# Visualizaciones

# Visualizaciones - Base

Opté por usar Tableau para hacer las visualizaciones. Los datos que me resultaron de interés fueron:

- Distribución de mis gastos. ¿Cuánto ahorro, y cuánto **no**?
- Por mes, ¿a qué le destino más dinero?
- ¿Hay alguna relación entre el precio del dólar y lo que ahorro?

El resultado está en una Story de Tableau.

# Visualización - Categorías

**Por mes, ¿a qué le destino más dinero?**

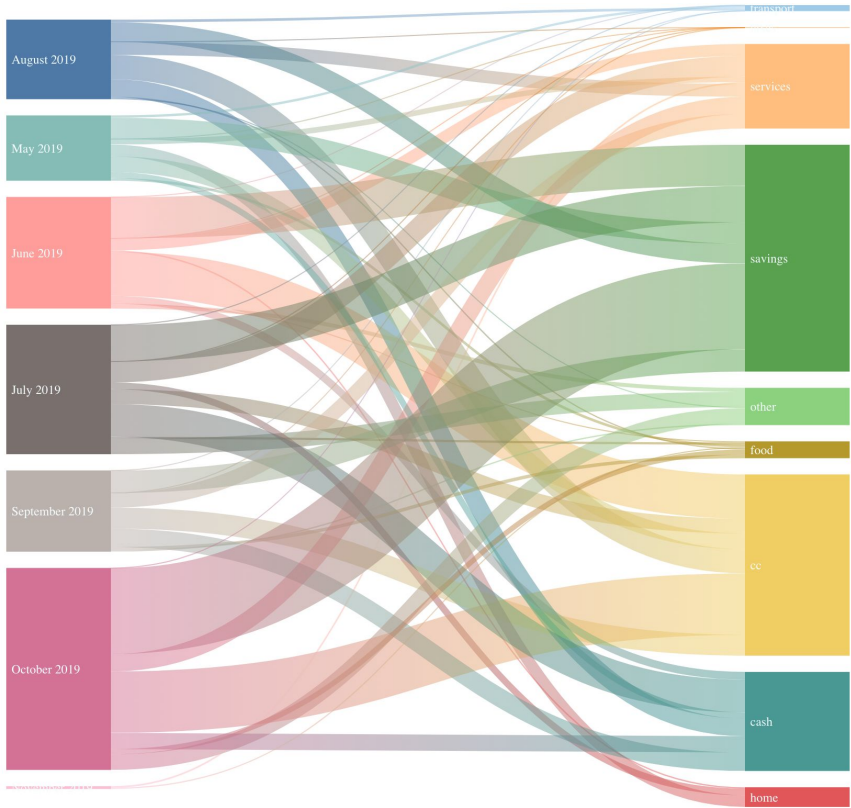
Usé un diagrama de Sankey. Para ello, aproveché un plugin de Tableau (luego de batallar horas y horas con hacerlo manualmente).

El diagrama permite seleccionar la dimensión de agregación según:

- Mes-Año
- Concepto
- Categoría

Finanzas 2019

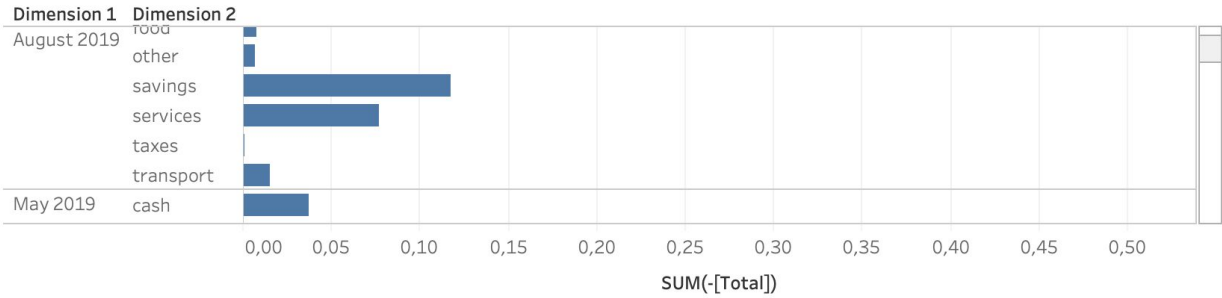
Cash Flow por mes	Dolar Scare: ahorro vs precio de venta	Destino del dinero	Factores Externos
-------------------	--	--------------------	-------------------



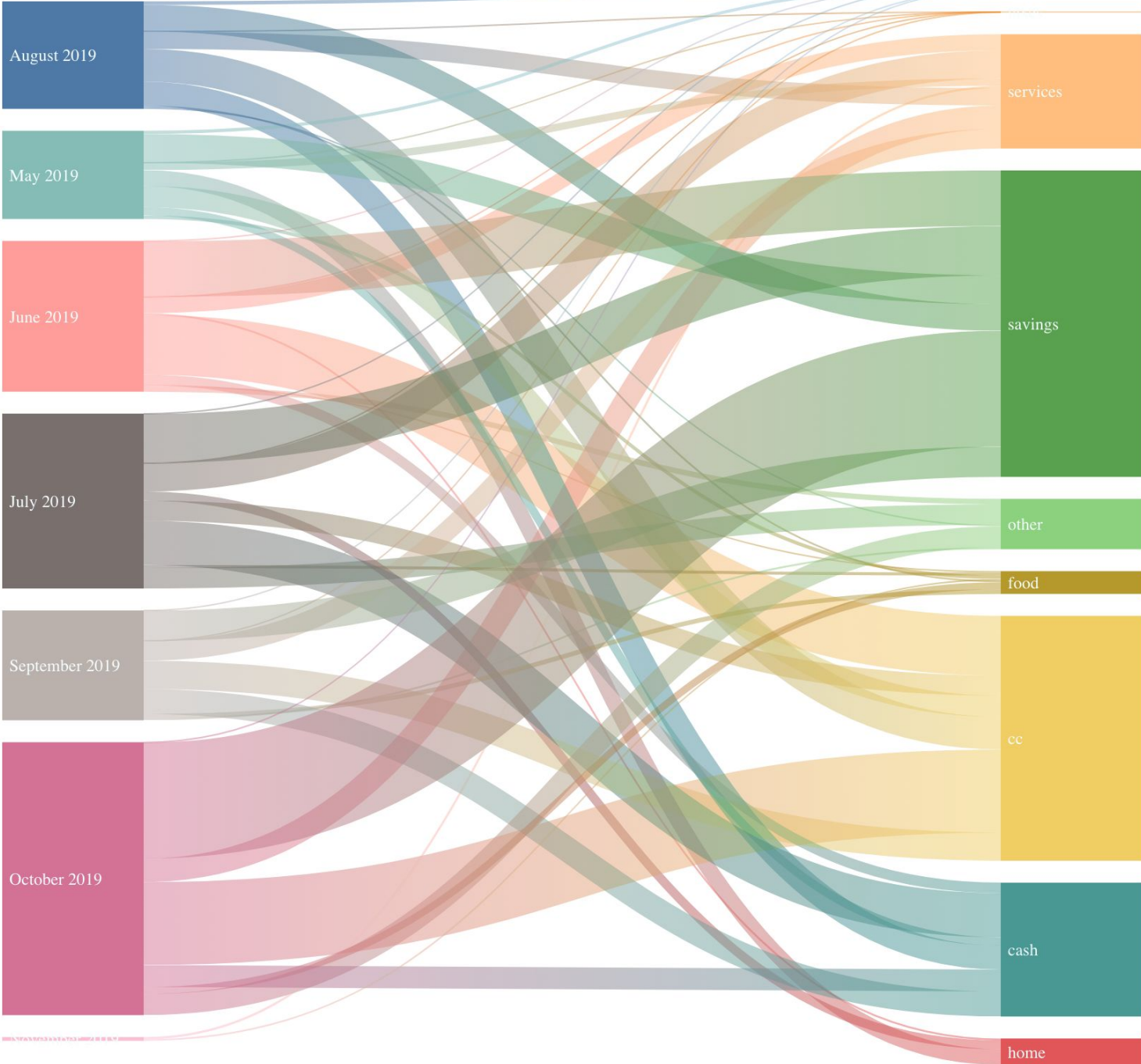
Select Dimension 1  
Mes

Select Dimension 2  
Concepto

Distribución de categorías y meses



## Vista detalle



# Visualización - Distribución

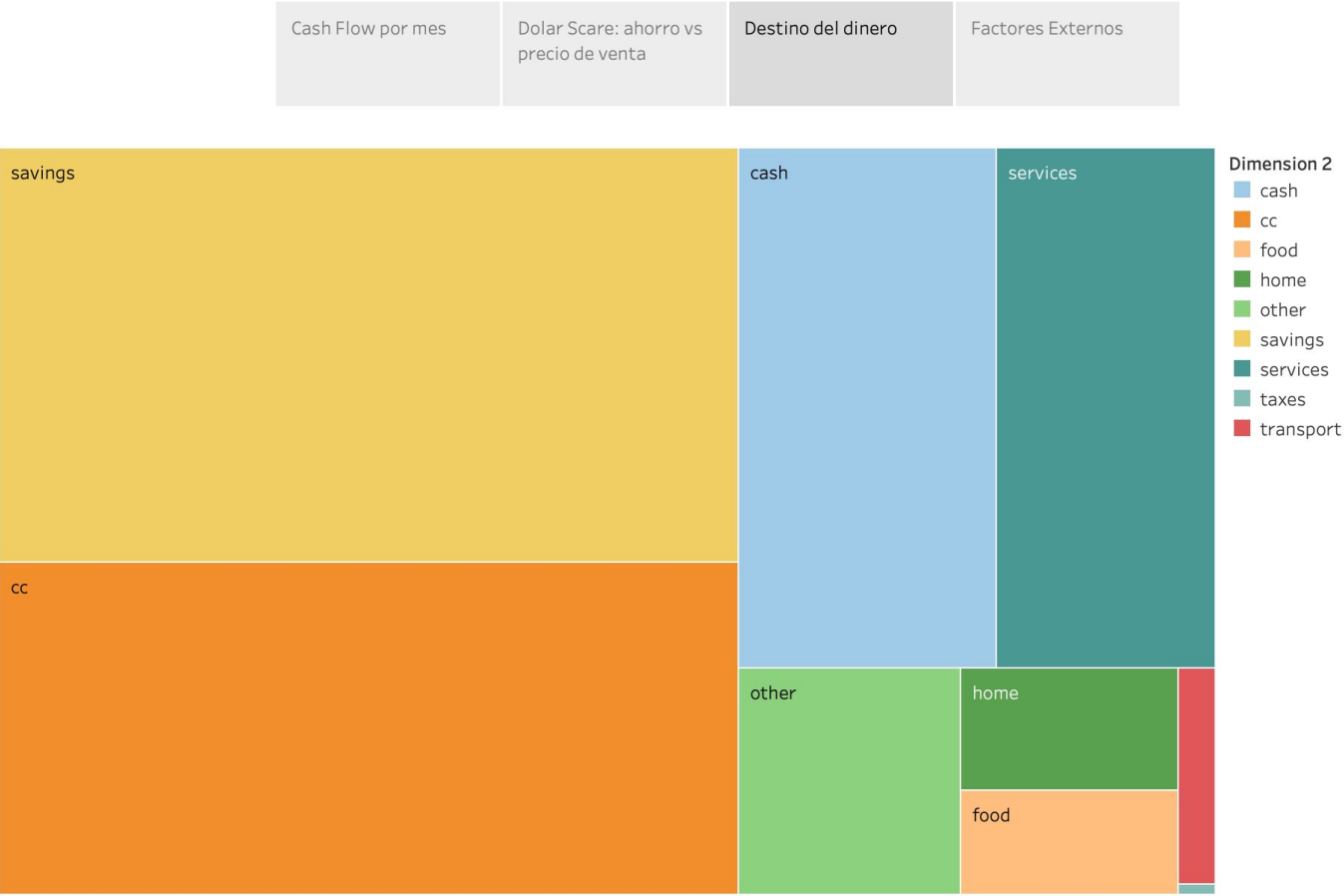
¿Cuánto llego a ahorrar?

¿En qué es lo que más gasto? ¿Comida, tarjeta, supermercado, servicios?

Opté por usar un treemap ordenado por las categorías. Esta data condensa el total en el tiempo entre ahorro, servicios y otros.



Finanzas 2019



Treemap de gastos

# Visualización - The Dollar Scare

Culturalmente, la divisa americana está muy presente.

¿Cuánto nos afecta?

(¿Cuánto **me** afecta al estrés?)

¿Qué decisiones nos fuerza a tomar su valor?

# Visualización - The Dollar Scare

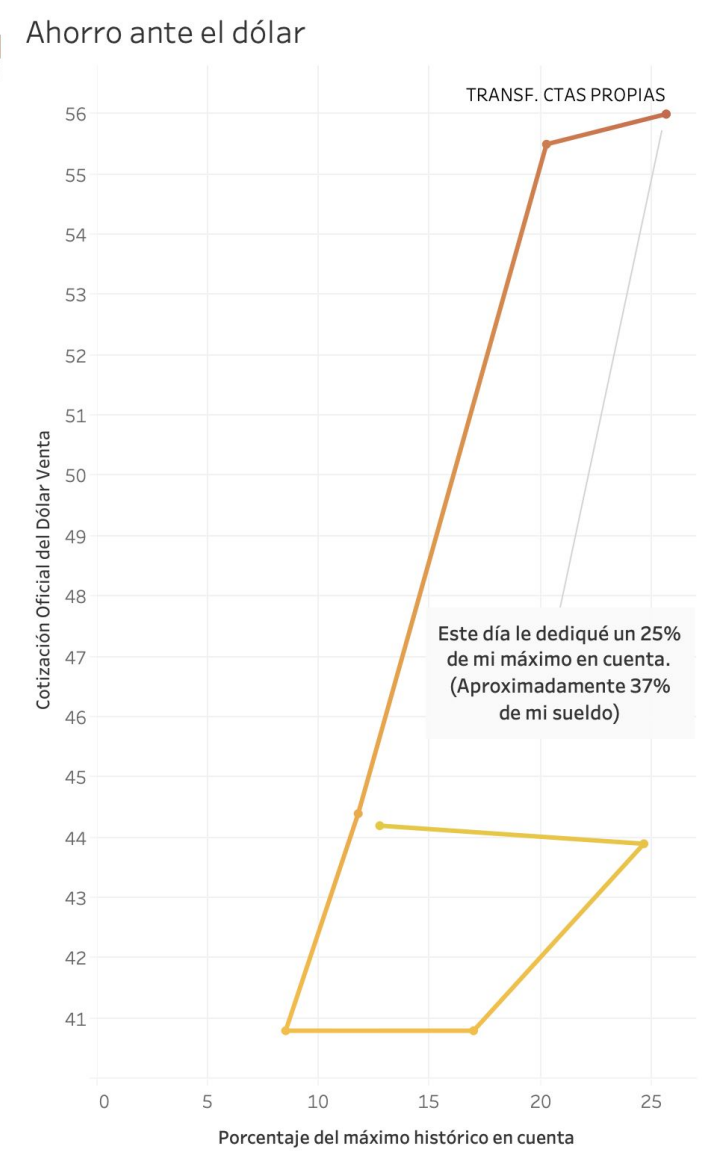
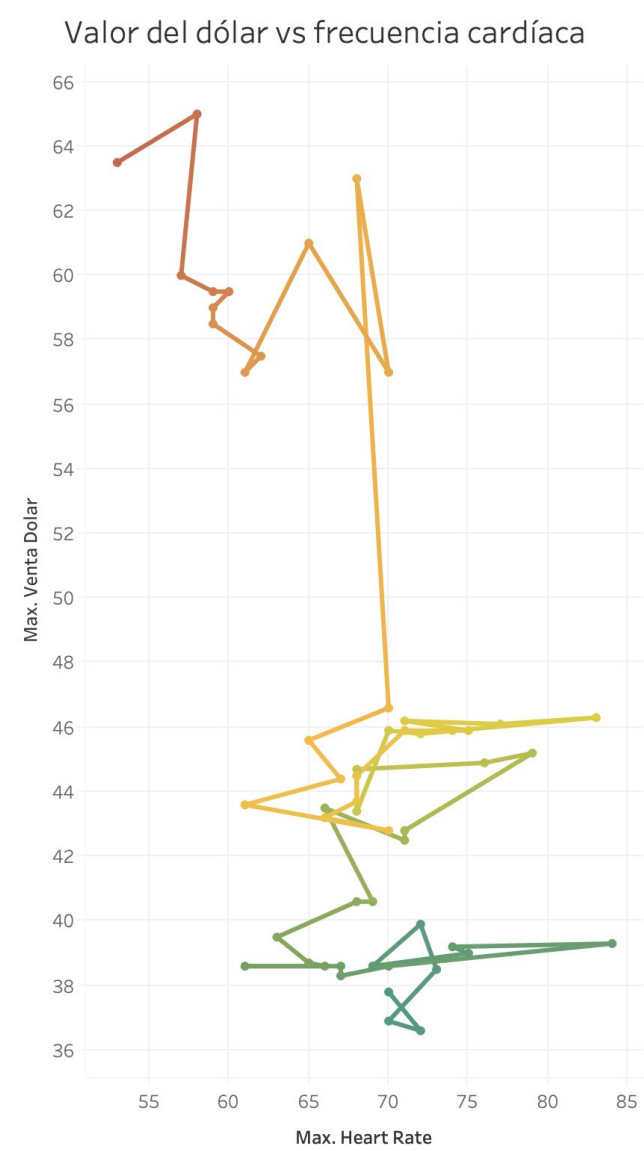
Plantee dos gráficos para comparar el comportamiento:

- Frecuencia cardíaca basal (indicadora de estrés) vs. valor de venta oficial
- Cantidad de compra vs. valor de venta oficial.

Para visualizarlo, usé un connected scatterplot que une con una línea temporal la comparación entre las dimensiones.

***Nota:*** el color indica proximidad a la fecha actual. Más rojo, más reciente.

Cash Flow por mes	Dolar Scare: ahorro vs precio de venta	Destino del dinero	Factores Externos
-------------------	--	--------------------	-------------------



# Visualización - Factores Externos

¿Puede haber alguna otra explicación para la frecuencia cardíaca?

¿Puede ser una correlación espuria?

Mi emprendimiento quedó hace poco seleccionado en el último batch de YCombinator (y recibió inversión). ¿Habrá algo de relación?

***Nota:*** la frecuencia cardíaca se marca con color. Verde es baja, roja es alta.

Finanzas 2019

Cash Flow por mes	Dolar Scare: ahorro vs precio de venta	Destino del dinero	Factores Externos
-------------------	--	--------------------	-------------------



# Conclusiones

# Conclusiones basadas en los gráficos

- El treemap es excelente para visualizar los principales cost drivers y poner en perspectiva las dimensiones de los números. No tiene en cuenta la inflación y variaciones de los ahorros, pero en la escala usada es seguro ignorarlos.
- El diagrama de Sankey es muy útil para entender el flow entre dos tipos de dimensiones: en las diapositivas aparece mes vs categoría. Puede compararse a mayor detalle agregando una columna más de "concepto". Lo interesante es que se puede "leer de ambos lados": de mes a categoría, o vice versa.



# Conclusiones basadas en los gráficos

- El connected scatterplot nos da una forma de comparar en un plano tres variables:
  - Frecuencia cardíaca (o montos destinados a ahorro)
  - Cotización del dólar
  - Tiempo / Fecha
- El linechart con clave de color nos da una forma rápida de entender una progresión temporal y de ubicar eventos sobre esa misma línea de tiempo.

# Conclusiones basadas en los datos

- La relación más fuerte entre la frecuencia cardíaca basal está dada por otros eventos y no por la cotización del dólar. :)
- ...Pero hay muy pocos datapoints para entender mi comportamiento de ahorro.
- La regla 50/30/20 dice que hay que guardar 50% para los gastos de comida, alquiler y salud, 30% para ocio y 20% en ahorro. El treemap nos dice que:
  - Ahorro es **33% :D**
  - Gastos clave: **45%**
  - Ocio: **15%**