
Viagens Federais 2017

— Clodoaldo e Renan —

Agenda

- A base
- Características dos dados
- Questões levantadas
- Filtragem dos dados
- Resultados
 - 1ª Visualização StreamGraph
 - 2ª Visualização WordCloud

Fonte da Database

Para a busca da base inicial de dados foi escolhido o site do Portal da Transparência como fonte de coleta dos dados com o foco sobre viagens realizadas a serviço, podendo este ser acessado através do seguinte endereço:

- <http://www.portaltransparencia.gov.br/download-de-dados/viagens>



A Database

Sobre o tema escolhido, foi decidido que os dados fossem referentes ao ano de 2017. A seguir é apresentado as características da base:

- Tamanho: 635 MB
- Formato: CSV (Comma-separated Values)
- Quantidade de arquivos: 4
- Título dos arquivos: Pagamento, Passagem, Trecho, Viagem

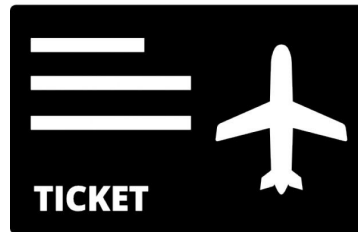


Características: Tabela Pagamento



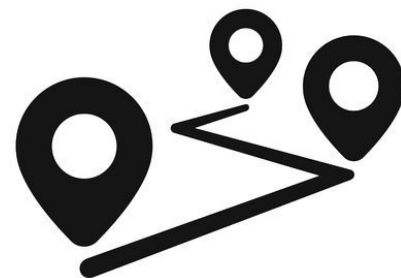
- Tamanho: 154 MB
- N° de Tuplas: 795.243
- Colunas:
- ID do processo de viagem, ID do órgão superior, Nome do órgão superior, Código do órgão pagador, Nome do órgão pagador, Código da unidade gestora pagadora, Nome da unidade gestora pagadora, Tipo de pagamento e Valor

Características: Tabela Passagem



- Tamanho: 59.80 MB
- N° de Tuplas: 273.622
- Colunas: ID do processo de viagem, Meio de transporte, País-Origem ida, UF-Origem ida, Cidade-Origem ida, País-Destino ida, UF-Destino ida, Cidade-Destino ida, País-Origem volta, UF-Origem volta, Cidade-Origem volta, País-Destino volta, UF-Destino volta, Cidade-Destino volta, Valor da passagem e Taxa de serviço

Características: Tabela Trechos



- Tamanho: 151.80 MB
- Nº de Tuplas: 1.038.852
- Colunas: ID do processo de viagem, Sequência Trecho, Origem-Data, Origem-País, Origem-UF, Origem-Cidade, Destino-Data, Destino-País, Destino-UF, Destino-Cidade, Meio de transporte, Número Diárias, Missão?

Características: Tabela Viagem



- Tamanho: 196.10 MB
- N° de Tuplas: 485.125
- Colunas: ID do processo de viagem, Situação, ID órgão superior, Nome do órgão superior, ID órgão solicitante, Nome órgão solicitante, CPF viajante, Nome, Cargo, Período Data de início, Período Data fim, Destinos, Motivo, Valor diárias, Valor passagens, Valor outros gastos

Filtragem dos dados

Visando a redução do tamanho da base de dados os seguintes passos foram adotados para a filtragem:

1. Importamos os dados para o MySQL
2. Calculamos a média dos gastos totais das viagens
3. Eliminamos os registros com valores abaixo da média
 - a. Valor médio obtido: 445.44 (R\$)

Resultados da Filtragem dos dados

- Diminuição geral no tamanho da base
- De 635MB para 132MB
- Tuplas antes do filtro: 481.592
- Tuplas depois do filtro: 131.375



Questões levantadas



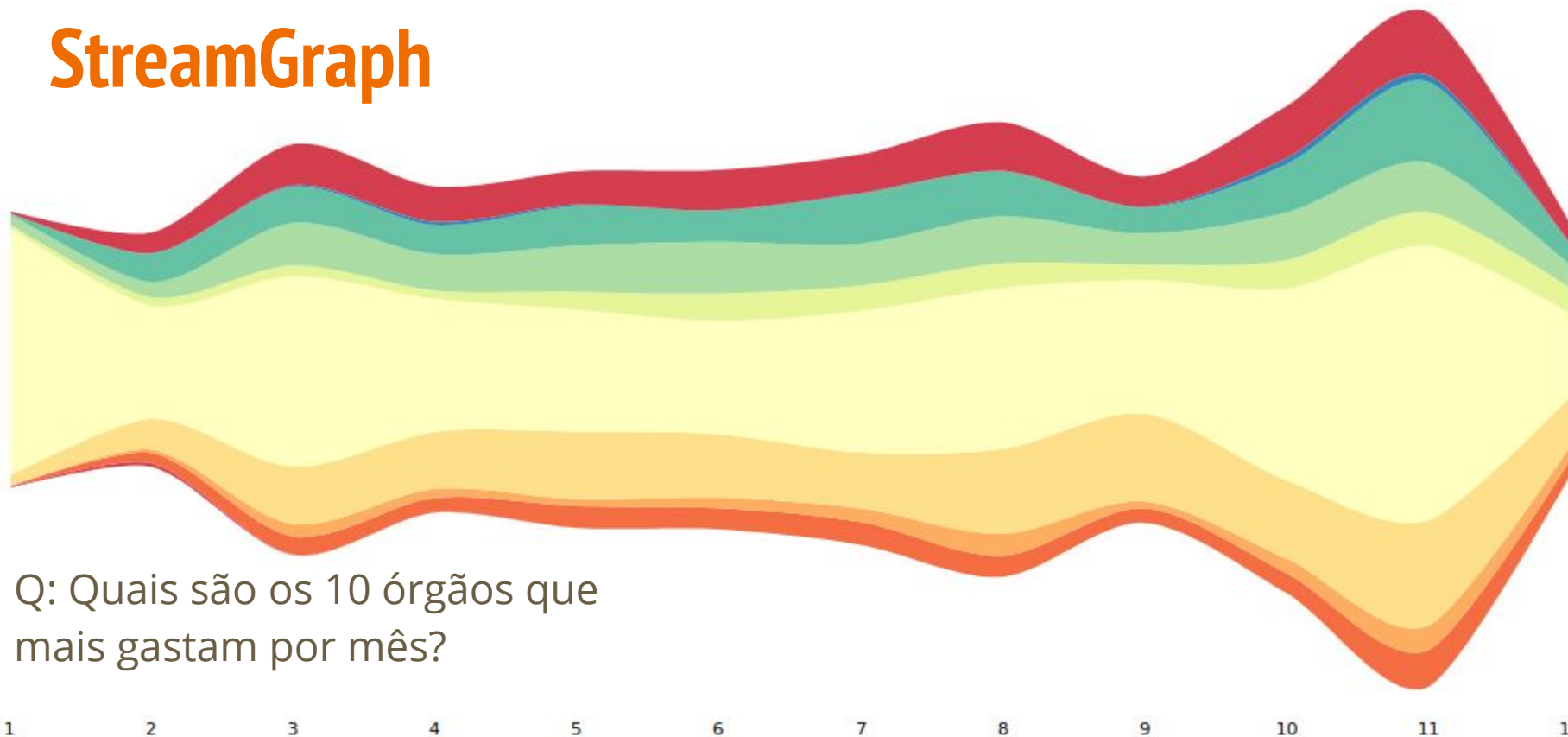
- 1. Quais são os 10 órgãos que mais gastam por mês?
- 2. Quais foram os principais motivos das viagens?

1ª- Visualização StreamGraph



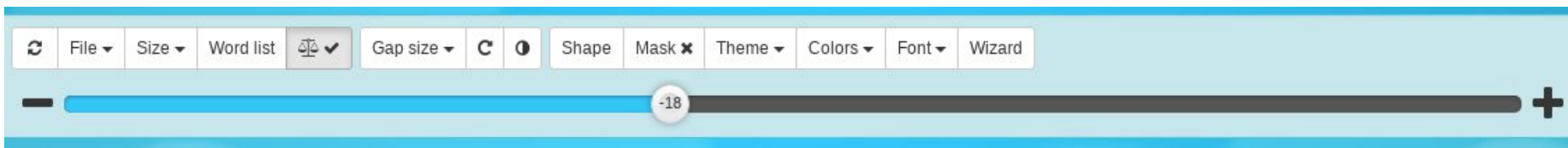
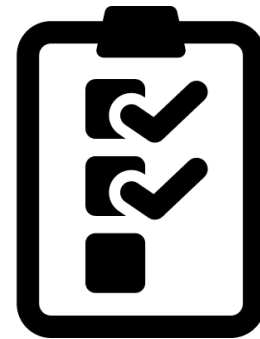
- Foi utilizado a linguagem R para a implementação;
- Bibliotecas utilizadas:
 - <Library> Streamgraph
 - <Tool> Devtools
- Foi necessário renomear órgão MINISTÉRIO DA JUSTIÇA E SEGURANÇA PÚBLICA para MINISTÉRIO DA JUSTIÇA.

StreamGraph



2ª- Visualização Cloud Words

- Remoção de stop Words
- Dados separado por meses
- Ferramenta disponível em: <https://www.wordclouds.com/>



Cloud Words (Janeiro e Fevereiro)



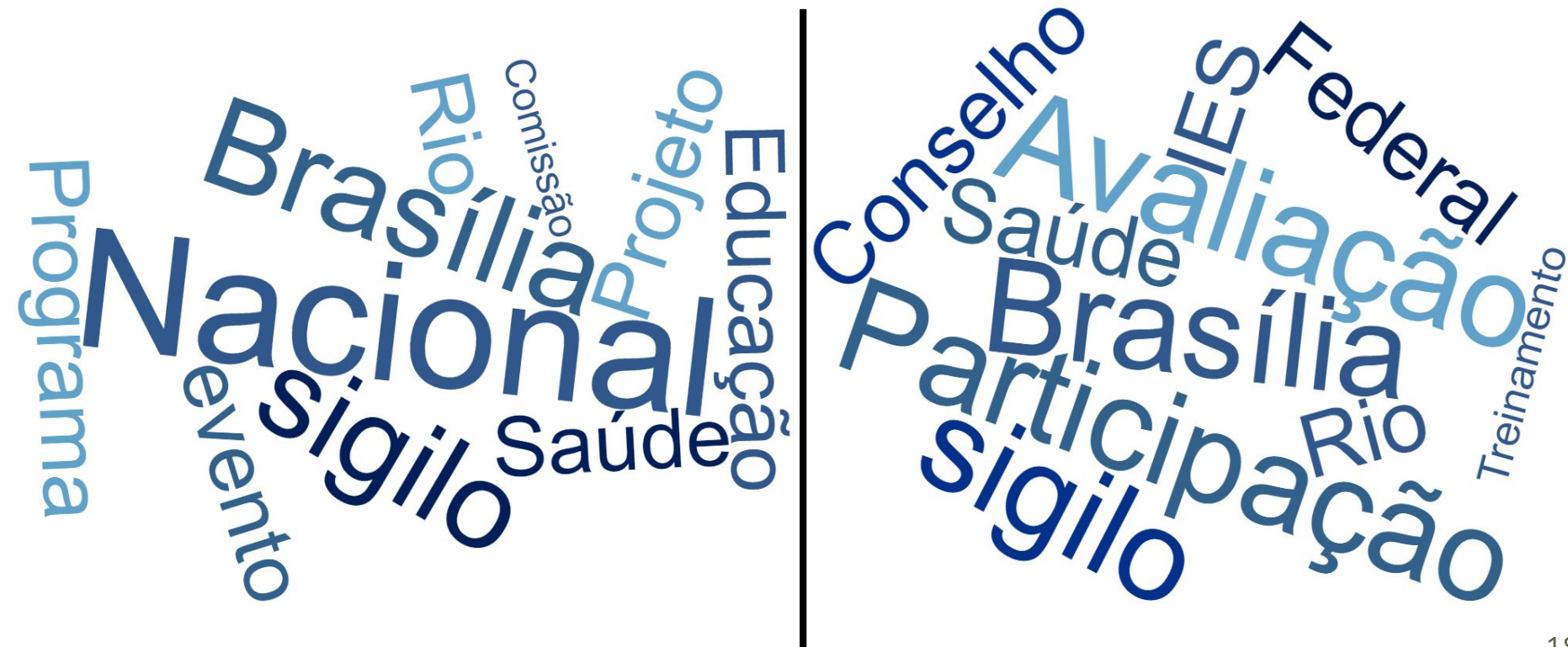
Cloud Words (Março e Abril)



Cloud Words (Maio e Junho)



Cloud Words (Julho e Agosto)



Cloud Words (Setembro e Outubro)



A word cloud for the months of September and October. The words are arranged in a circular pattern around the center. The most prominent words are 'Período' and 'Julho'. Other visible words include 'Internacional', 'Rio', 'Ensino', 'IES', 'humanos', 'SUS', and 'Provisória'.

Período
Julho
Internacional
Rio
Ensino
IES
humanos
SUS
Provisória



A word cloud for the months of September and October. The words are arranged in a circular pattern around the center. The most prominent words are 'Nacional' and 'sigilo'. Other visible words include 'Avaliação', 'Educação', 'IES', 'Rio', 'Federal', 'Saúde', and 'reunião'.

Nacional
sigilo
Avaliação
Educação
IES
Rio
Federal
Saúde
reunião

Cloud Words (Novembro e Dezembro)



3ª Visualização Choropleth

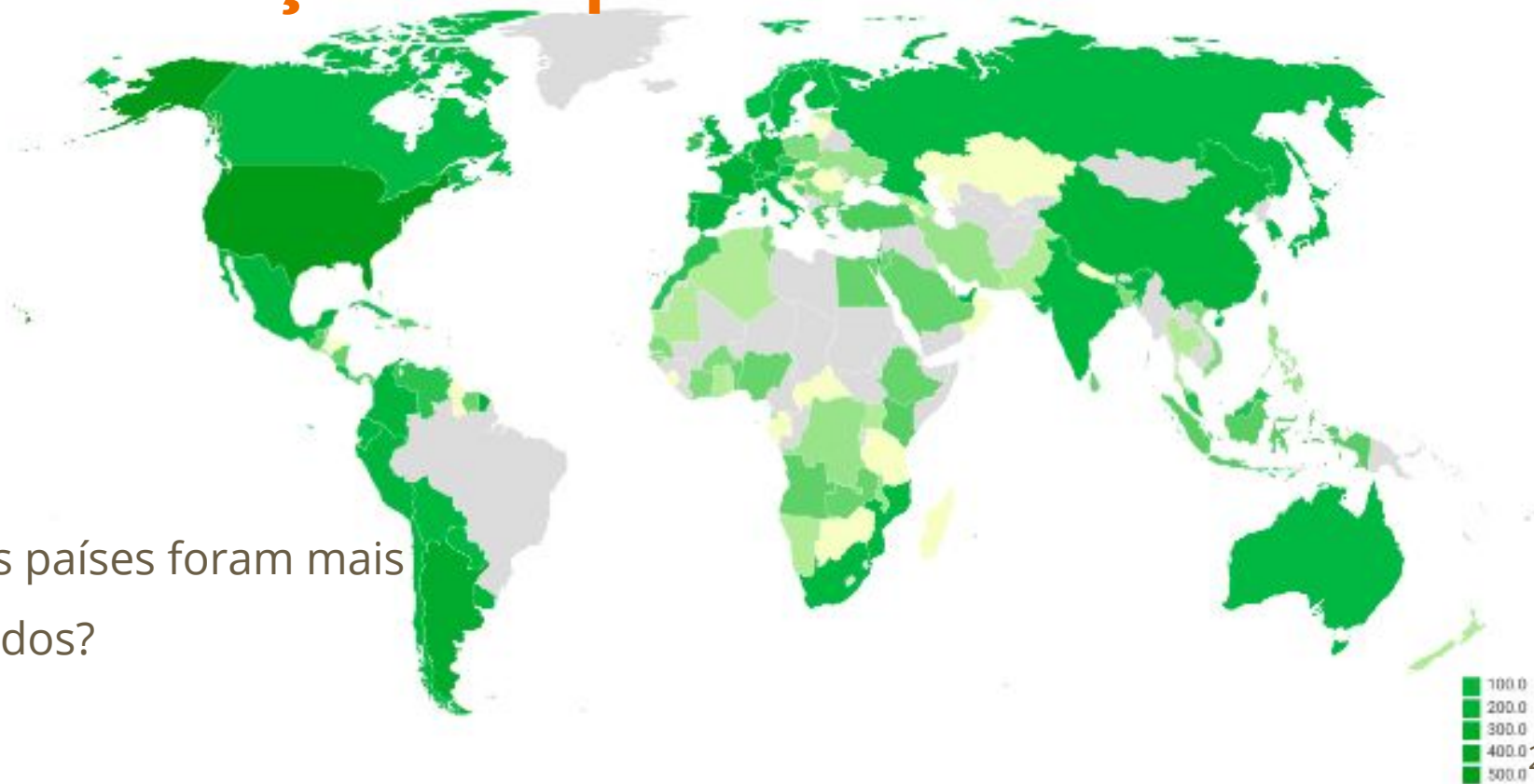
- Países que não foram incluídos:
 - French guiana com 10 visitas
 - Hong kong 11 visitas
 - Ferramenta utilizada:

<https://app.datawrapper.de/map/llWvy/basemap>

3ª Visualização Choropleth

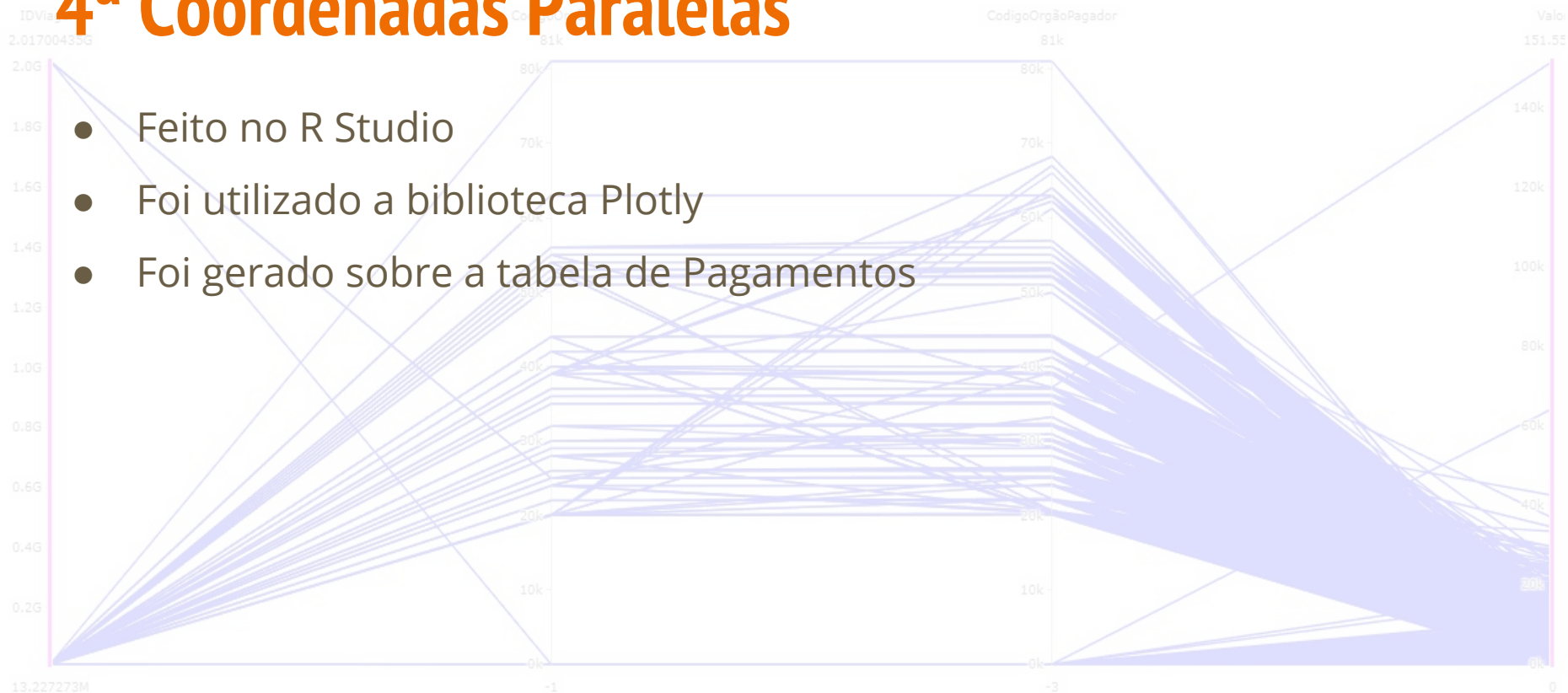


- Quais países foram mais visitados?

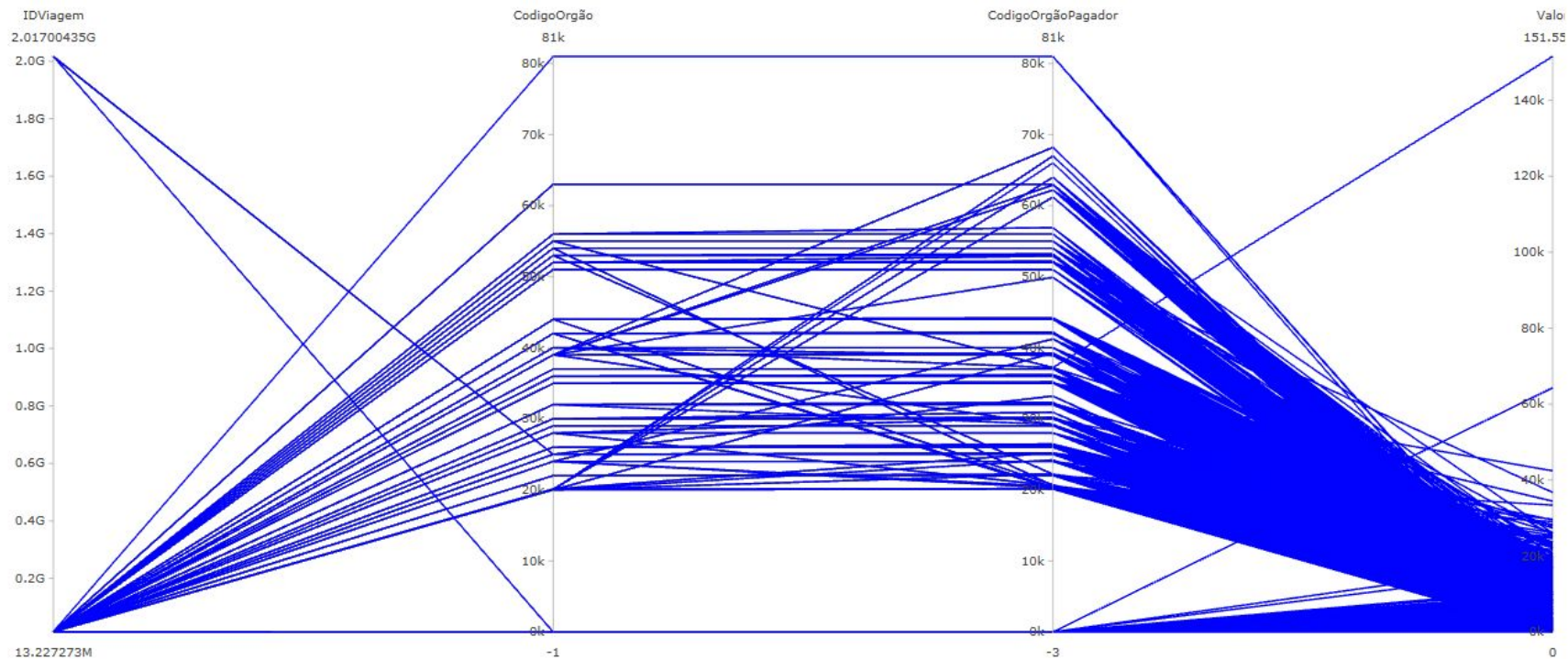


4ª Coordenadas Paralelas

- Feito no R Studio
- Foi utilizado a biblioteca Plotly
- Foi gerado sobre a tabela de Pagamentos



4ª Coordenadas Paralelas



Dúvidas?

