



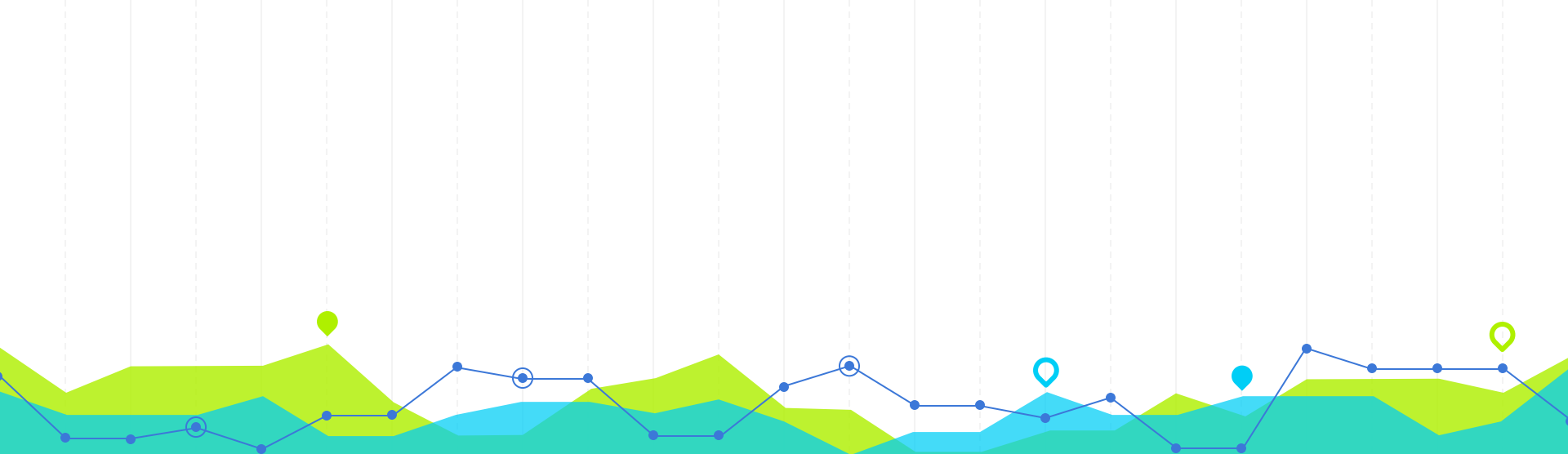
Manipulando Bases de Dados Públicas em R

Iago Rios

Membro do Mandacaru Hackerspace, desenvolvedor e professor

iagorssantos@gmail.com | linkedin.com/in/iagorios



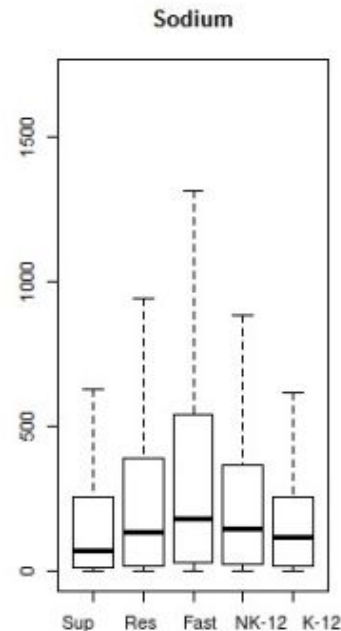
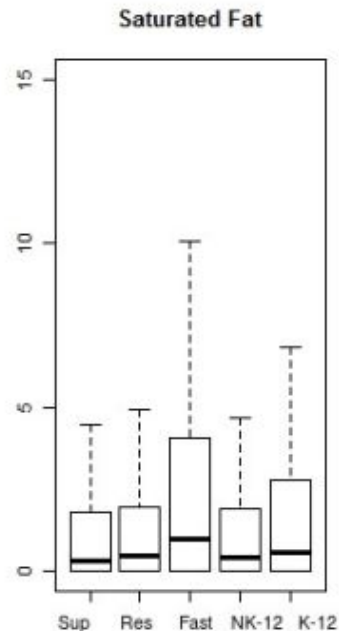
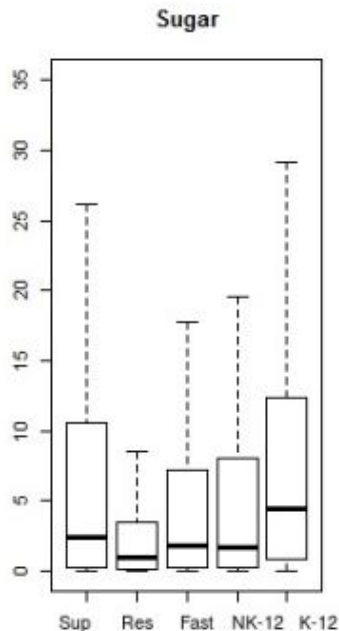


Experiências Anteriores

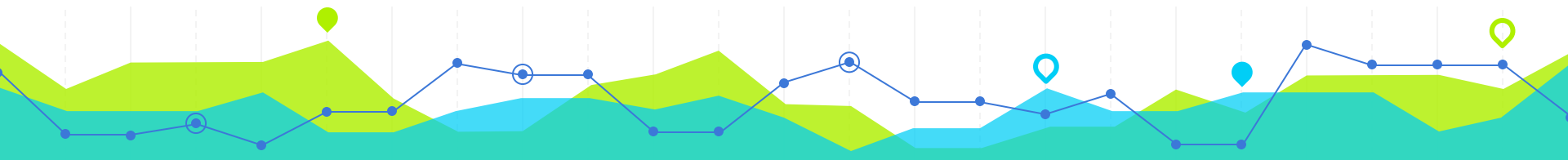
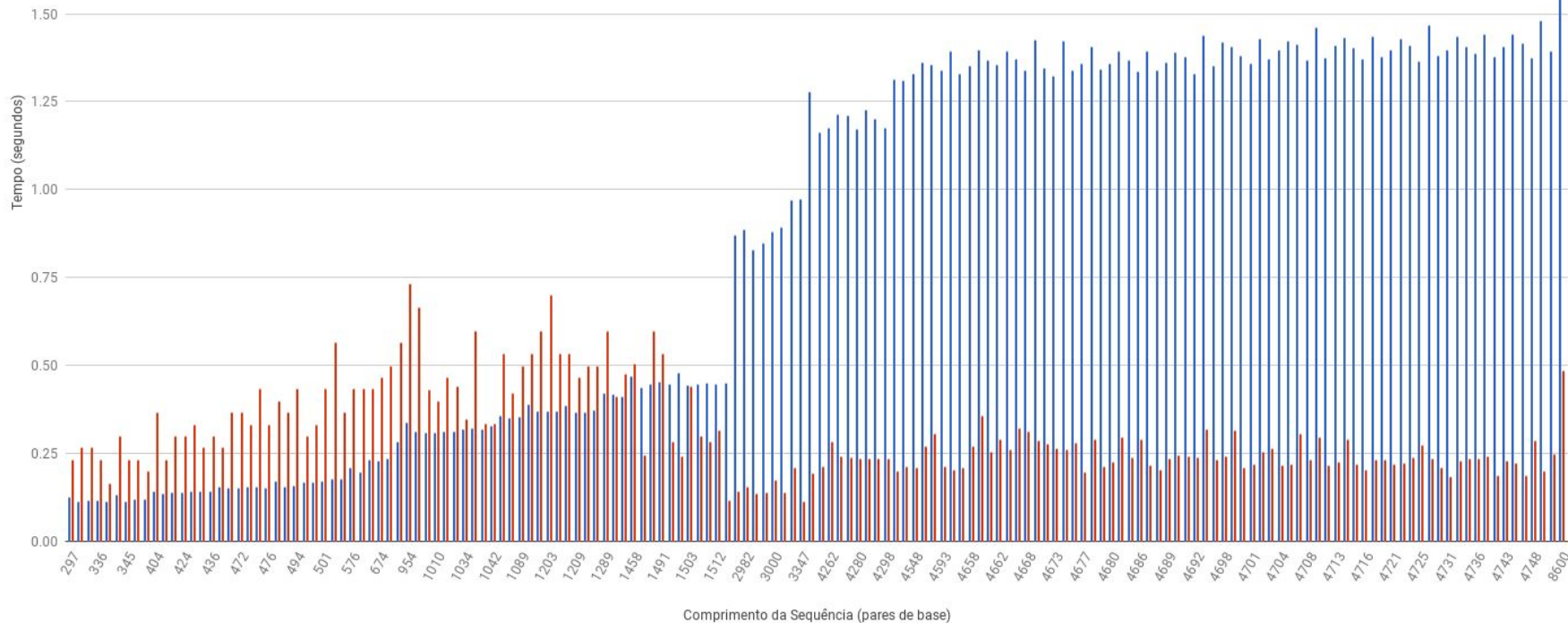
Alguns trabalhos com grande volume de dados

0

Consumo de Açúcar em Cafeterias de Escolas Americanas

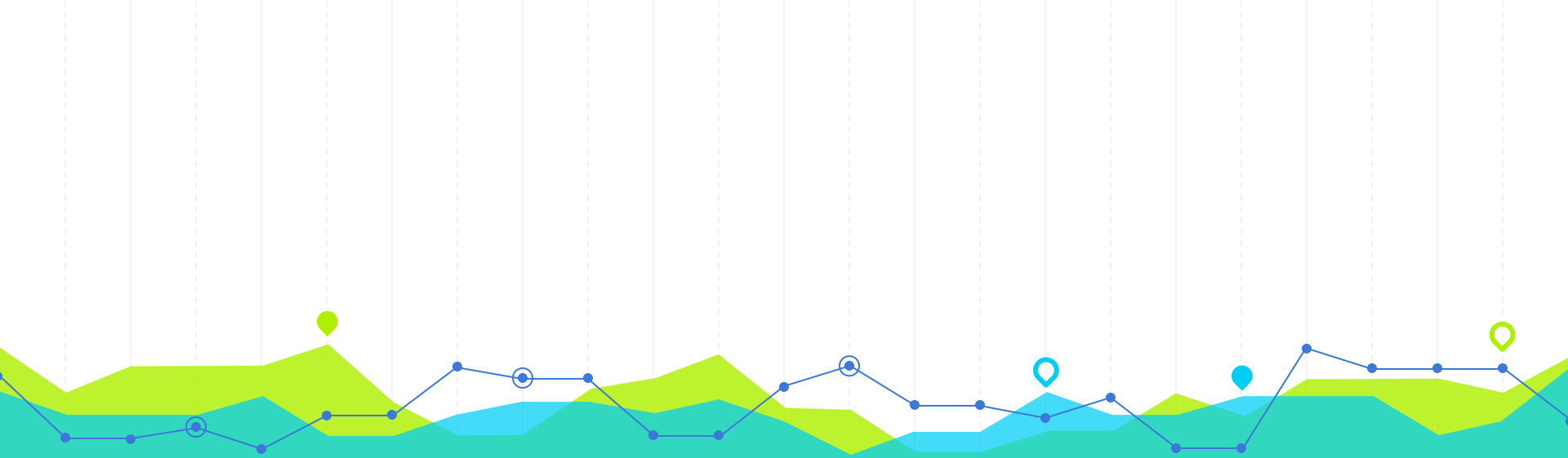


Análise de sequências do vírus HIV



Rendimento de Alunos do Fundamental 1 em Matemática





Introdução

Dados no Brasil e no mundo

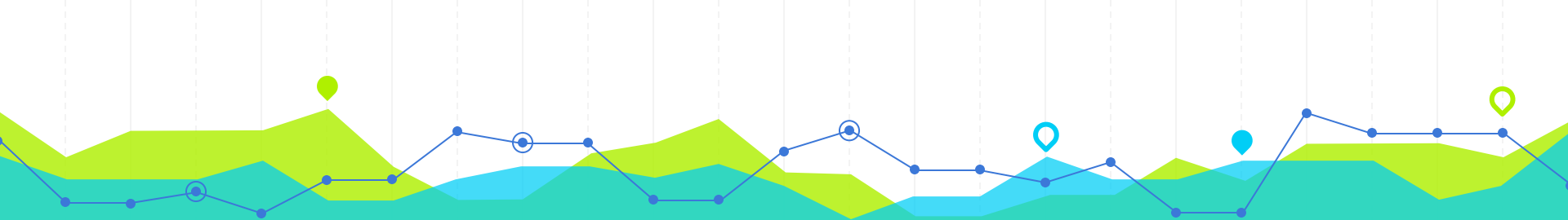
1

“

*90% dos dados foram gerados nos
últimos 2 anos*

Science Daily, 2013





2,7 zettabytes

De dados atualmente



Em 18 de novembro de 2011 foi sancionada a Lei de Acesso a Informação Pública (Lei 12.527/2011) que regula o acesso a dados e informações detidas pelo governo

Portal Brasileiro de Dados Aberto



OGP Global Summit 2018: Tbilisi

LEARN MORE



PORTAL BRASILEIRO DE DADOS ABERTOS

foi um dos primeiros compromissos firmados no
plano de ação de governo aberto





PORTAL BRASILEIRO DE DADOS ABERTOS



[Dados](#) | [Organizações](#) | [Aplicativos](#) | [Inventários](#) | [Concursos](#) | [INDA](#) | [Perguntas frequentes](#) | [Contato](#) | [Sobre o portal](#)

PESQUISAR

em 5443 conjuntos de dados com 28201 recursos (o que é isto?)

Conjuntos de dados em destaque

Conjuntos de dados colocados em destaque pela equipe do dados.gov.br.

Preços de Medicamentos

Os dados representam a lista de preços de Medicamentos , contemplando o preço Fábrica, ou preço fabricante (PF), que é o preço praticado...

Opendata AIG Brazil - Ocorrências Aeronáuticas na Aviação Civil...

![Modelo de Dados]

(https://raw.githubusercontent.com/nosbielcs/opendata_aig_brazil/master/reference/data)

Publicações mais recentes

| Conjunto de dados | Data | Responsável |
|-------------------|------------|---|
| CONTRATOS | 19/05/2018 | Instituto Federal de Educação, Ciênc... |
| COMPRAS DIRETAS | 19/05/2018 | Instituto Federal de Educação, Ciênc... |
| LICITAÇÕES | 19/05/2018 | Instituto Federal de Educação, Ciênc... |
| Contratos | 18/05/2018 | Universidade Federal de Roraima - UF... |
| Licitações | 18/05/2018 | Universidade Federal de Roraima - UF... |

A maior plataforma de visualização de dados sociais e econômicos do Brasil. Totalmente aberta e livre.

O DataViva é uma ferramenta de pesquisa que disponibiliza dados oficiais sobre exportações, atividade econômica, localidade, educação e ocupações de todo o Brasil. São 11 aplicativos, que, juntos, formam mais de 1 bilhão de possibilidades de visualizações.

Inicie uma pesquisa >

Assista ao vídeo Institucional

EMPREENDEDORES

Identifique o perfil econômico e as oportunidades de negócios de uma região

[Veja depoimentos](#)

AGENTES DE DESENVOLVIMENTO

Avalie a criação de políticas de desenvolvimento de acordo com a localidade

[Veja depoimentos](#)

ESTUDANTES E PROFISSIONAIS

Descubra informações sobre empregos disponíveis, renda por ocupação e cursos

[Veja depoimentos](#)

Explore nossos bancos de dados

O DataViva reúne dados de diversas fontes: Emprego e Renda (RAIS), Comércio Exterior Brasileiro (SECEX/MDCI), Educação Básica e Ensino Técnico (Censo Escolar), Ensino Superior (Censo da Educação Superior) e Comércio Internacional (UN Comtrade).

O que você gostaria de descobrir?

Localidades Brasileiras



Ocupações



Atividades Econômicas



Produtos



Parceiros Comerciais



Universidades



Ensino Superior



Curso Básico



O Brasil em dados libertos

Repositório de dados públicos disponibilizados em formato acessível



Coleção de CNPJs e CPFs brasileiros

Documentos coletados de dados públicos, a partir dos seguintes datasets: socios-brasil, gastos-diretos, gastos-deputados, eleicoes. Nota: os CPFs foram ofuscados por questões de privacidade.

[VISUALIZAR](#)

Cursos e notas de corte do PROUNI 2018

Quantidade de vagas, mensalidade e notas de corte do PROUNI 2018.

[VISUALIZAR](#)

Eleições Brasil

Dados das eleições brasileiras, desde 1996, capturados do TSE, como: filiados a partidos, candidatos e votação.

[VISUALIZAR](#)

Empresas sócias de outras empresas

Empresas sócias de outras empresas

[VISUALIZAR](#)

Estado da balneabilidade da costa baiana

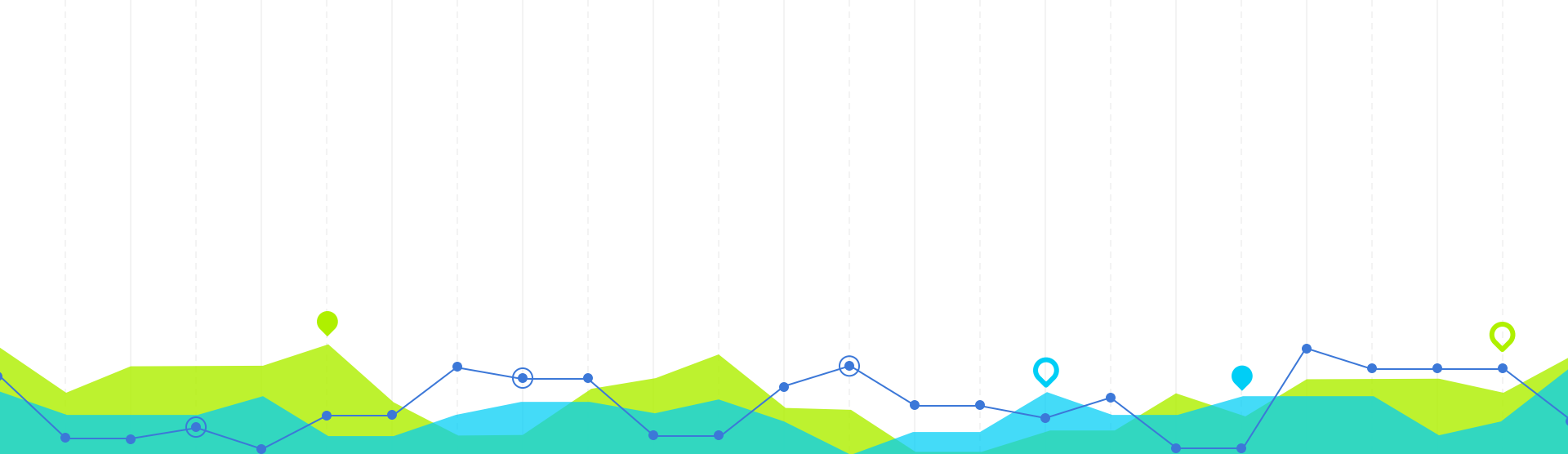
Condição de balneabilidade das praias da costa baiana, retirados dos boletins semanais do INEMA.

[VISUALIZAR](#)

Filiações a Partidos Políticos

Lista completa de filiações a partidos políticos brasileiros, disponibilizada pelo TSE. Estão incompletos os dados de candidatos e de resultados das eleições de 1994 a 2002.


[VISUALIZAR](#)



Motivação

Ciência de dados, necessidade ou modinha?

2

A night cityscape with a complex network of glowing blue lines overlaid, symbolizing data and information. The lines form a dense web connecting various points across the image, with some lines appearing as bright streaks of light. In the background, city buildings are visible, including one with a 'UBS' sign. The overall scene conveys a sense of high-tech connectivity and data flow.

É MUITA INFORMAÇÃO

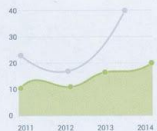


Morris Charts

Line Chart



Area Chart



Bar Chart

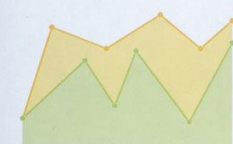


Donut Chart



Sparkline Charts

Line Chart



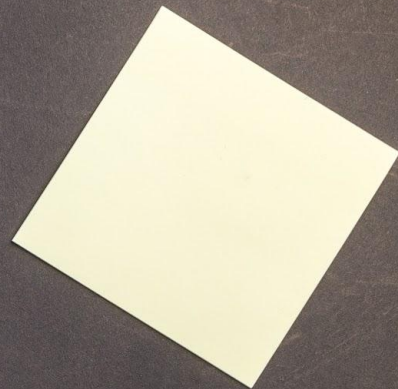
Bar Chart



Pie Chart



Easy Pie Charts





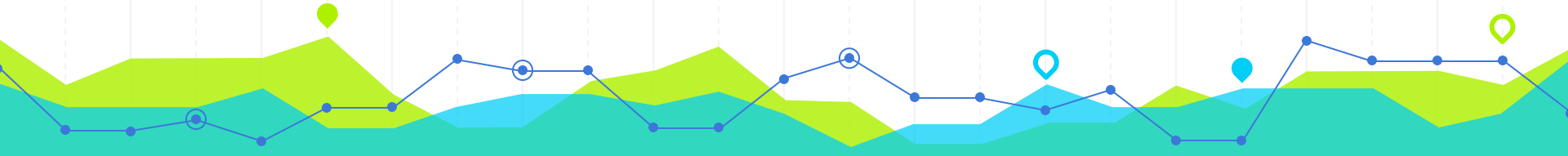
Papillonair
Summer 3
Standard
Our Price \$549

Exemplo: negociação com bikes

Ideação

Pesquisa de Mercado

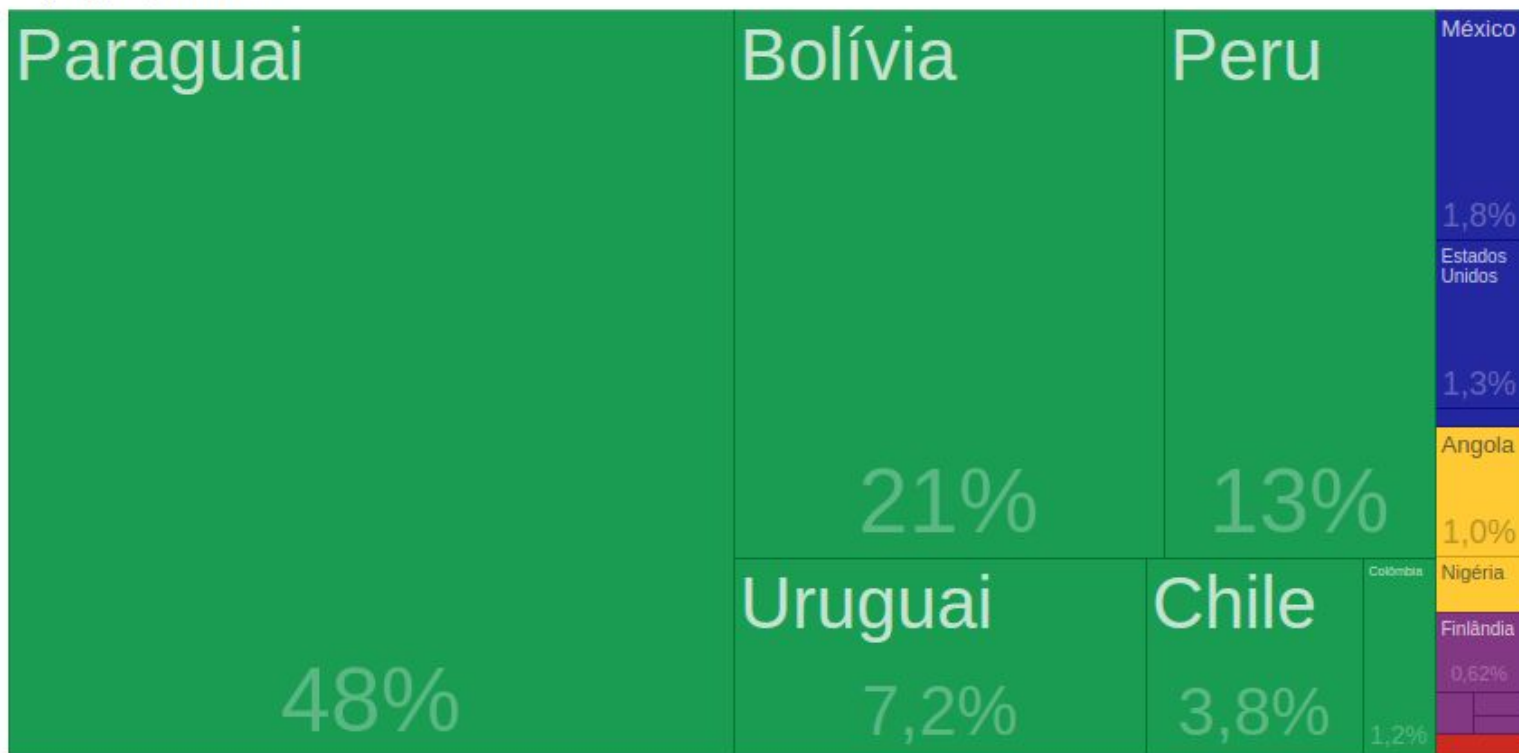
Implementação



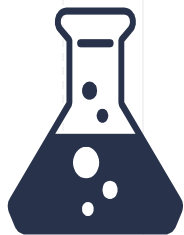
Origens das Importações/Destinos das Exportações das Bicicletas do Brasil (2015)

Baseado nos Estados Produtores

Exportações: \$394 Mil USD



Veja mais visualizações na versão completa do site DataViva.info.



Outras áreas...

A Ciência de Dados pode impactar outras áreas
das nossas vidas?



RESEARCH ARTICLE

Automatic Prediction of Cardiovascular and Cerebrovascular Events Using Heart Rate Variability Analysis

Paolo Melillo^{1,2*}, Raffaele Izzo³, Ada Orrico^{1,2}, Paolo Scala², Marcella Attanasio^{1,2}, Marco Mirra³, Nicola De Luca^{3*}, Leandro Pecchia⁴

1 Multidisciplinary Department of Medical, Surgical and Dental Sciences, Second University of Naples, Naples, Italy, **2** SHARE Project, Italian Ministry of Education, Scientific Research and University, Rome, Italy, **3** Department of Translational Medical Sciences, University of Naples Federico II, Naples, Italy, **4** School of Engineering, University of Warwick, Coventry, United Kingdom

* paolomelillo85@gmail.com (PM); nideluca@unina.it (NDL)



HANDS-ON



Como responder as nossas questões com dados?

Escolher uma base de dados

Preparação

Manipulação





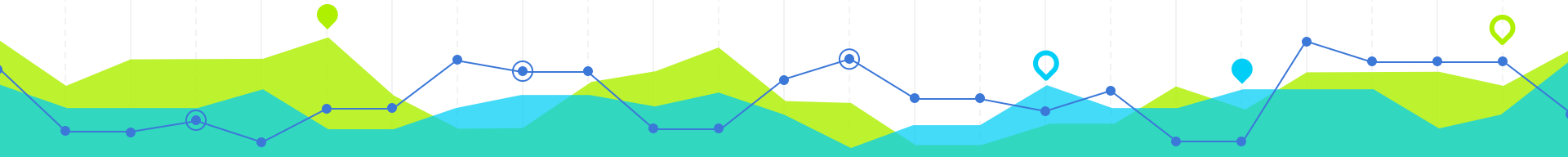
A Nossa Base de Dados

Vamos dar uma olhada em dados da educação

3

Download:

brasil.io/dataset/cursos-prouni



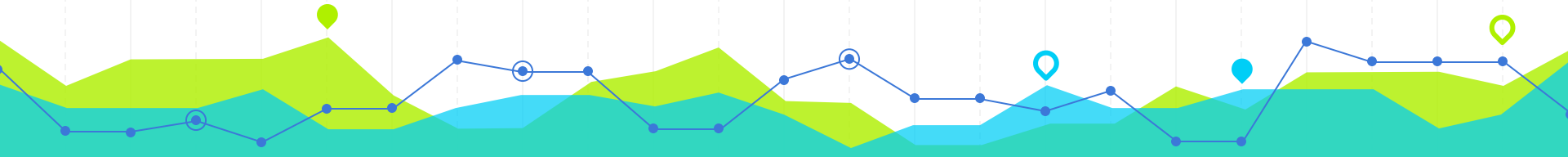


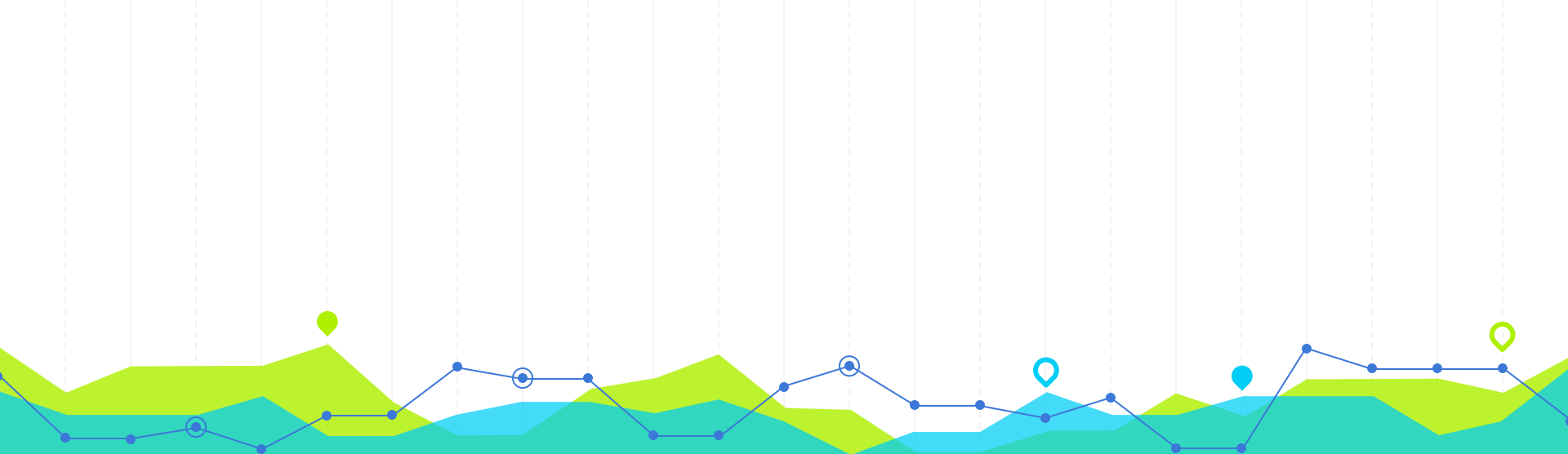
Visualização dos Dados e Primeiras Ideias

Sim, vale usar o Excel também

4

O que podemos extrair desse dataset?





Preparação da Base de Dados

Parte muito importante para o sucesso do trabalho

5

Limpeza dos Dados

- Remoção (ou não) de outliers
- Normalização de strings e de variáveis
- Formatos de datas
- Valores ausentes
- Inconsistências óbvias

É muito importante. Principalmente para o avanço da análise





Manipulação da Base com R

Agora é a parte boa :)

6

Sobre a Linguagem

- Uma linguagem para fins estatísticos e análise de dados
- Atualmente ocupa o 11º lugar no TIOBE index
- Código aberto
- Desenvolvida em C, Fortran e R



Formas de Utilização

- Terminal
- RStudio
- R Commander



print("Obrigado!")

Iago Rios

iagorssantos@gmail.com | [linkedin.com/in/iagorios](https://www.linkedin.com/in/iagorios)

