

PROJETO 1

ANÁLISE EXPLORATÓRIA DE DADOS

PERFIL DOS BRASILEIROS AO LONGO DE ALGUNS ANOS

PNAD:

<https://ww2.ibge.gov.br/home/estatistica/populacao/trabalhoerendimento/pnad2015/default.shtm>

A Pesquisa Nacional por Amostra de Domicílios - PNAD investiga anualmente, de forma permanente, características gerais da população, de educação, trabalho, rendimento e habitação e outras, com periodicidade variável, de acordo com as necessidades de informação para o País, como as características sobre migração, fecundidade, nupcialidade, saúde, segurança alimentar, entre outros temas. O levantamento dessas estatísticas constitui, ao longo dos 49 anos de realização da pesquisa, um importante instrumento para formulação, validação e avaliação de políticas orientadas para o desenvolvimento socioeconômico e a melhoria das condições de vida no Brasil.

A PNAD considera que uma **unidade domiciliar** é o domicílio particular ou uma unidade de habitação (quarto, apartamento etc.) em domicílio coletivo, podendo esse estar localizado em área urbana ou rural. Já a **população residente** foi composta pelos moradores presentes e ausentes, ou seja, pelas pessoas que tinham a unidade domiciliar (domicílio particular ou unidade de habitação em domicílio coletivo) como local de residência habitual e, na data da entrevista, estavam presentes ou ausentes, temporariamente, por período não superior a 12 meses em relação àquela data.

A base de dados resultante dessa pesquisa é oferecida no site do IBGE em dois formatos: arquivo de Pessoas – em que cada linha descreve as características das pessoas residentes em um domicílio; e o arquivo de Domicílios – em que cada linha descreve as características particulares de um domicílio.

Nesse projeto, iremos trabalhar com o arquivo de **PESSOAS**.

PERFIL DOS BRASILEIROS AO LONGO DE ALGUNS ANOS

PROBLEMA:

Esse projeto tem como proposta levantar possíveis *semelhanças, diferença e/ou melhorias* em algumas possíveis vertentes:

- Descrever perfil **das mulheres como chefes de família**; ou
- Descrever perfil **dos aposentados brasileiros**; ou
- Descrever perfil **dos brasileiros economicamente ativos**; ou
- Descrever perfil **dos estudantes brasileiros de diversas escolaridades ou de uma em específico**; ou
- Descrever **outro perfil desde que antes discutido com a professora**.

Assim, você deverá:

1. Escolher **APENAS UMA DAS OPÇÕES** acima para ser o seu projeto;
2. Descrever um **OBJETIVO AO SEU PROBLEMA** independente de qual seja a opção escolhida;
3. **ESCOLHER E/OU CRIAR VARIÁVEIS** que julgue importante de acordo com seu objetivo;
4. Analisar o perfil escolhido dentro das diversas variáveis escolhidas que julgar necessárias para auxiliar nessa compreensão.
5. Apenas para UMA característica, selecionar o mesmo perfil **da PNAD de 2011 e comparar com o obtido em 2015. IDENTIFICAR UMA EVOLUÇÃO OU NÃO AO LONGO DO TEMPO** dessa particular característica estudada.

Obrigatoriamente, sua análise deverá considerar **um misto de variáveis quantitativas e qualitativas** e usar as diversas ferramentas estatísticas que serão vistas na disciplina.

Importante! Se preocupe não só em fazer um filtro das variáveis que julgar serem relevantes para diferenciar as pessoas do perfil escolhido (colunas do *dataframe*), mas também avaliar se há a necessidade ou não de ser feito um filtro das pessoas que deseja analisar (linhas do *dataframe*). Por exemplo, se deseja analisar trabalho de crianças e adolescentes até 14 anos, deve antes excluir todas as linhas referentes às pessoas com idade acima de 14 anos; e assim por diante. A importância do “filtro de linhas” é focar exclusivamente nas pessoas que possuem as características que deseja analisar. Ainda, é importante que se dedique a entender o problema e traçar um objetivo claro para que possa ser capaz de selecionar e trabalhar com as

variáveis mais adequadas. A utilização do dicionário será extremamente importante para localizar o nome e significado de cada variável considerada em cada ano de pesquisa.

Por fim, escolha as ferramentas estatísticas mais adequadas para explorar nas suas análises. Seu projeto deve conter um *mix* entre tabela de frequências, gráfico de setor, gráfico de barras, histograma e boxplot, além de medidas resumo. É válido ressaltar que a escolha dessas ferramentas deve ser norteada pela adequação aos tipos de variáveis e ao questionamento do problema. Cada recurso estatístico utilizado no seu projeto deve deixar sempre claro o que se deseja investigar e qual a interpretação que se faz dos resultados obtidos.

Links de algumas notícias do IBGE que podem motivar um perfil dos brasileiros que deseja investigar:

<http://g1.globo.com/tecnologia/noticia/2016/11/brasil-supera-marca-de-100-milhoes-de-internautas-diz-ibge.html>

<https://g1.globo.com/df/distrito-federal/noticia/df-e-a-unidade-da-federacao-que-mais-utiliza-internet-no-brasil-diz-ibge.ghtml>

<https://g1.globo.com/economia/noticia/trabalho-infantil-quase-1-milhao-de-menores-trabalham-em-situacao-ilegal-no-brasil-aponta-ibge.ghtml>

<http://g1.globo.com/economia/noticia/2016/11/nivel-de-escolaridade-dos-pais-influencia-rendimento-dos-filhos.html>

<https://g1.globo.com/economia/noticia/diferenca-de-salario-medio-de-homens-e-mulheres-pode-chegar-a-quase-r-1-mil-no-pais-aponta-ibge.ghtml>

Entre várias outras matérias que pode encontrar no site <http://g1.globo.com/>

BASE DE DADOS (LINKS PARA DOWNLOAD):

PNAD 2015

DADOS: <https://www.dropbox.com/s/9vncz5fn37m5iid/PES2015csv.zip?dl=0>

DICIONÁRIO: https://www.dropbox.com/s/nwm1sn7g3504kdm/DicionarioVariaveisPessoas_PNAD2015.xls?dl=0

PNAD 2011 (USAR APENAS PARA UMA CARACTERISTICA)

DADOS: <https://www.dropbox.com/s/5glbdemuf6vbdnf/PES2011csv.zip?dl=0>

DICIONÁRIO: https://www.dropbox.com/s/prjm8tgnkl85zdp/DicionarioVariaveisPessoas_PNAD2011.xls?dl=0

REGRAS:

1. O Projeto 1 é em DUPLA.
2. Elaborar um relatório que contenha as análises que contribuíram para as conclusões de acordo com o objetivo inicial. Use o **arquivo layout** disponibilizado na pasta Projeto1 do Blackboard.
3. O relatório deve ser feito em IPython Notebook (com células de texto formatadas em Markdown) e deve conter, pelo menos, as seguintes seções:
 - a. Introdução,
 - b. Filtro da base de dados,
 - c. Análises e explicação dos resultados,
 - d. Conclusão.

A estrutura do documento deve ser clara e de fácil compreensão da linha de raciocínio. Nesse caso, o notebook não deve haver excesso de impressões não discutidas de variáveis e de *dataframe*.

Aconselhamos fazer uma análise geral e, após finalizada, salve com outro nome, limpe seu IPython Notebook apenas com os resultados relevantes e melhore seu texto.

4. Seu projeto deve ser adicionado no seu github dentro de uma pasta chamada Projeto1.
5. O arquivo DEVE ser com extensão .ipynb

CRONOGRAMA:

DATA	Finalização:
14/02	APS1: Check descrito na Aula 02 (COM NINJAS ATÉ 20/02)
21/02	APS2: Check descrito na Aula 04 (COM NINJAS ATÉ 25/02)
01/03	PROJETO 1 FINALIZADO (COMBINAR COM ALUNOS) Fazer git push em seu Github até 23:59 do Projeto 1 finalizado.

RUBRICS DE AVALIAÇÃO DO OBJETIVO DE APRENDIZADO

Objetivo de aprendizado	Insatisfatório (I)	Em desenvolvimento (D)	Essencial (C)	Proficiente (B)	Avançado (A)
Elaborar análises exploratórias de dados (univariadas e multivariadas), utilizando ferramentas estatísticas e computacionais adequadas.	<p>Não consegue trabalhar com bases de dados de forma proficiente.</p> <p>Apresenta problemas com arquivos, formatos de arquivos ou não tem habilidades básicas de filtragem de dados.</p>	<p>Consegue abrir e filtrar dados de forma básica, mas não consegue conectar os dados à prática de análise.</p> <p>O notebook é praticamente uma coleção de comandos do panda e gráficos, com pouca análise e interpretação.</p> <p>Tem dificuldade para lidar com valores inválidos e não consegue eliminá-los, corrompendo a análise.</p>	<p>Descreve com riqueza de detalhes o objetivo que deseja avaliar nesse projeto de acordo com a vertente escolhida.</p> <p>Constrói uma base de dados com variáveis pertinentes ao objetivo definido, sejam essas do tipo qualitativo e/ou quantitativo.</p> <p>É capaz de filtrá-las, cruzá-las e processá-las para embasar sua análise.</p>	<p>Executa os comportamentos da rubrica C de maneira excepcional.</p> <p>Analisa ocorrências conjuntas com recursos adequados ao tipo de variável (crosstab, gráficos de barras, de setor, histograma e/ou boxplot), respeitando o tipo de variável.</p> <p>Interpreta corretamente o que as ocorrências conjuntas significam em função dos objetivos da pesquisa.</p> <p>Comunica os resultados com bastante clareza.</p>	<p>Executa os comportamentos da rubrica B de maneira excepcional.</p> <p>Incorpora aspecto histórico, analisando todas as PNADs envolvidas na vertente escolhida.</p> <p>Expõe os resultados obtidos de maneira excepcional.</p>