

Instruções para o trabalho

1. Todos os grupos recebem os mesmos dados
2. Os arquivos estão em formato de texto (txt ou csv) ou xls.
3. O arquivo Dados_de_Publicacoes contém os seguintes cabeçalhos:
 1. id (de cada autor)
 2. id_institution (da instituição)
 3. year
 4. journal
 5. sub-discipline
 6. number_of_authors_publication
 7. weight_authorship (indica que a publicação é dividida pelo número de autores)
 8. weight_journal_indexation (indica em quantas sub-disciplinas está indexada a revista)
 9. weight_final (da divisão da publicação pelo número de autores e subdisciplinas indexadas.
4. O arquivo Instituições.txt contém:
 1. InstitutionID
 2. InstitutionName
 3. State_ID
 4. Mesoregion_ID
 5. Microregion_ID
 6. Municipality_ID
 7. Municipality_Name
5. O arquivo Subdisciplines.xls contém os códigos das subdisciplinas assim como os nomes.
(corresponde a 554 nomes de campos de estudo)
6. O arquivo PIB dos Municípios - base de dados 2010-2015.xls contém os dados socioeconômicos das prefeituras (municipalities), destacando entre elas: PIB, PIB per capita, População por ano (entre 2010 e 2015)
7. Os arquivos RAIS_N_ESTABELECIMENTOS_xxxx (por ano), contém as informações das atividades econômicas desenvolvidas (contadas pelo número de estabelecimentos) em cada prefeitura por ano.
8. As diretrizes do trabalho são:

0. Os Dados_de_Publicacoes, apresenta em cada linha um trabalho (artigo) de um pesquisador, indicando que vocês devem agrupar por instituição ao qual o pesquisador está associado. Vejam que algumas das instituições apresentam um número pequeno de trabalhos ao longo do período (vocês podem decidir de acordo com algum critério, um ponto de corte).

1. juntar todas as informações possíveis (produção científica e atividade econômica)

2. Extrair a maior quantidade de conclusões possíveis com respeito da interação ou influência da atividade científica sobre a atividade econômica.

3. Estudar a posição das instituições nas principais localidades, podendo agrupar os dados de acordo com micro região ou mesoregião (para as atividades econômicas) ou por estado para as atividades científicas.

4. Quanto melhor o mapeamento entre as informações, maior será a nota.

9. O formato de entrega (EM R e PYTHON), será utilizando a livreria (Rmarkdown) do r, ou similar em Python, de forma que o relatório final seja carregado junto com os programas. O formato do relatório deve ser o de um relatório final com introdução, metodologia, e deve conter uma análise dos resultados. Esta requisição é para que possa analisar a saída do programa (relatório).

10. A data de entrega será o dia 20 de dezembro, pelo e_disciplinas ou pelo e-mail tuesta@usp.br