

## Ficha 2

### Wrangling

1. Abra o ficheiro **NOTAS.Rdata** e transforme os *data frames* **NOTAS**, **NOTAS2** e **NOTAS3** em formato *tidy*, utilizando os comandos *gather*, *spread* e/ou *separate*.

NOTAS1			
aluno	Periodo1	Periodo2	Periodo3
João	7.6	7.9	8.3
Pedro	6.1	6.9	7.4
Bruno	7.4	8.0	8.2
Ivo	8.8	8.5	7.5
Francisco	7.3	7.7	8.3
Raquel	6.8	7.3	7.6

NOTAS2		
aluno	variavel	valor
João	idade	22
João	nota	2
Pedro	idade	21
Pedro	nota	5
Amanda	idade	18
Amanda	nota	10
Fernando	idade	19
Fernando	nota	12

NOTAS3	
aluno	Notas P1/P2/P3
João	7.6/8/8.9
Pedro	10/12/12.9
Bruno	9/8.2/11.9
Ivo	12.6/11/12.9

2. Considerando o ficheiro **Insc2017\_2018.xls**, com informação relativa ao número de alunos inscritos no ensino superior (DGEEC), **indique**:
- 2.1. O número de estabelecimentos existentes no ficheiro;
  - 2.2. O número de cursos de “Arquitectura” existentes na “Área Metropolitana de Lisboa”;
  - 2.3. **Por estabelecimento**:
    - i) o número de alunos inscritos, por sexo;
    - ii) a percentagem de alunos estrangeiros inscritos. Liste, em primeiro lugar, os estabelecimentos com maior percentagem de alunos estrangeiros;
    - iii) os cursos existentes na área geral “Artes e Humanidades”. Ordene esta listagem pelos estabelecimentos com mais oferta nesta área.

3. Considere os ficheiros **estat\_servicos.xls** e **CAE\_rev3.xlsx**. **Importe-os** para o ambiente R. Tendo por base a variável **CAE3** (Classificação Portuguesa de Atividades Económicas, Revisão 3) presente no ficheiro **estat\_servicos.xls**:

i) Crie um *data frame* (**Estat\_DIV**) com informação sobre o número de empresas (**N\_EMP**) e o valor médio de VVN por DIVISÃO, listando primeiro as divisões com mais empresas.

CAE3	DIVISÃO
52220	52

- ii) Ao *data frame* obtido anteriormente, adicione uma coluna com a descrição da Divisão (esta última informação presente no ficheiro **CAE\_rev3.xlsx**);
- iii) Calcule o número de empresas cuja atividade económica inclua a palavra “transportes”.

4. Importe o ficheiro **quadro\_ganhomedio.xls**, com informação do ganho médio mensal por localização geográfica (NUTS). Utilizando as ferramentas incluídas no *package dplyr*, calcule o valor médio dos ganhos dos 7 anos por NUTSIII (3 dígitos), ordenando o resultado por ordem decrescente de valor médio.

Localização	ValorMédio(2011_2017)
Alto Minho	€€€.€€
...	

5. Considere o ficheiro **Propostas.xlsx** que contém um histórico das propostas aprovadas no SIGUA (Sistema Integrado de Gestão de Universos e Amostras), para uma determinada operação estatística.

5.1 Indique, **por utilizador**:

- o **número total** de propostas efetuadas;
- o **mês** com mais propostas executadas;
- o **dia** em que foram efectuadas mais propostas de alteração.

5.2 Para cada utilizador, indique a percentagem de propostas executadas, durante o período da manhã e durante o período da tarde.