

Folha 2: Manipulação de Dados

Importação de Dados

1. O site [UCI ML Repository - Post-Operative Patient](#) contém um conjunto de dados com a informação sobre para onde devem ser enviados pacientes na área de recobro pós-operatório: cuidados intensivos, área geral do hospital, ou para casa.

- (a) Faça download do ficheiro [post-operative.data](#).
- (b) Importe o conjunto de dados para um data frame do R, garantindo que os valores desconhecidos são corretamente traduzidos para a nomenclatura do R.

2. No ficheiro [monthly-reported-number-of-chick.xlsx](#) estão registados o número de casos de varicela reportados mensalmente em Nova York entre 1931 e 1972.

- (a) Faça download do ficheiro e, usando a função `read_excel` da package `readxl` importe o conjunto de dados para um data frame do R.
- (b) Chame a função `plot` sobre o data frame criado.
- (c) Este conjunto de dados representa uma série temporal. Use a função `xts` da package `xts` para criar uma série temporal com valores indicados pela segunda coluna e datas indicadas pela primeira coluna.
- (d) Chame, novamente, a função `plot` mas agora sobre o objeto criado.
- (e) Produza o gráfico correspondente ao período a partir de 1970.
- (f) Produza o gráfico correspondente ao período entre 1970 e 1971.

Manipulação de Dados com dplyr

3. Crie um *data frame table* a partir do conjunto de dados `Aids2` da package `MASS`. Sobre este conjunto de dados, execute as operações necessárias de forma a responder às seguintes questões.

- (a) Inspecione as primeiras e as últimas linhas do dataset.
- (b) Mostre apenas as colunas cujo nome começa por `s`.
- (c) Mostre apenas as linhas referentes a pacientes do sexo feminino que estejam vivos.

- (d) Ordene as linhas pela idade e pela categoria de transmissão do vírus.
- (e) Crie, para os pacientes que morreram, uma nova coluna com a diferença em dias entre o momento do diagnóstico e o momento da morte.
- (f) Obtenha, para os pacientes que morreram, o número médio de dias de vida para cada categoria de transmissão.
- (g) Sobre o data set anterior obtenha, também para cada categoria, o número de pacientes total.
- (h) Adicione o sexo como outro nível de agrupamento e volte a executar a operação anterior.