

Importação de dados

Análise de dados com linguagem R

```
dadosHerica <- read.csv("Herika/Base_dados_Trabalho_Final.csv",sep = ";")
summary(dadosHerica)
```

```
##          UF          Populacao2015      Populacao2014      Populacao2013
## AC       : 1      12.286.176: 1      12.226.346: 1      12.163.338: 1
## AL       : 1      1.515.126 : 1      1.496.889 : 1      1.478.164 : 1
## AM       : 1      1.768.204 : 1      1.748.531 : 1      1.728.214 : 1
## AP       : 1      188.268   : 1      185.546   : 1      182.772   : 1
## BA       : 1      1.918.622 : 1      1.906.665 : 1      1.894.084 : 1
## CE       : 1      22.429.379: 1      22.195.749: 1      21.956.629: 1
## (Other):21      (Other)   :21      (Other)    :21      (Other)    :21
##          Populacao2012      Populacao2011      Populacao2010      PIB2015
## 11.467.083: 1      11.406.621: 1      11.347.502: 1      11.579.809 : 1
## 1.417.694 : 1      1.400.892 : 1      1.383.453 : 1      1.234.497 : 1
## 1.590.011 : 1      1.576.455 : 1      1.560.501 : 1      12.746.146 : 1
## 173.100   : 1      170.779   : 1      168.021   : 1      132.202.280: 1
## 1.826.813 : 1      1.815.563 : 1      1.803.642 : 1      1.387.243 : 1
## 21.108.679: 1      20.898.199: 1      20.680.319: 1      1.412.838 : 1
## (Other)   :21      (Other)   :21      (Other)   :21      (Other)   :21
##          PIB2014          PIB2013          PIB2012          PIB2011
## 10.747.541 : 1      10.752.239: 1      1.112.854 : 1      10.513.494: 1
## 1.216.806 : 1      1.105.353 : 1      1.123.318 : 1      1.054.662 : 1
## 12.419.736 : 1      11.523.116: 1      11.276.096: 1      1.085.819 : 1
## 1.336.246 : 1      1.157.332 : 1      1.190.392 : 1      1.125.045 : 1
## 1.378.974 : 1      1.350.836 : 1      135.692.170: 1      1.213.923 : 1
## 145.386.560: 1      14.076.924: 1      14.059.895 : 1      12.686.307: 1
## (Other)   :21      (Other)   :21      (Other)   :21      (Other)   :21
##          PIB2010          SMP2010          SMP2011          SMP2012
## 1.116.500 : 1      11.339.341: 1      10.917.240: 1      10.284.331: 1
## 11.652.080 : 1      12.304.739: 1      13.188.743: 1      11.831.680: 1
## 120.204.650: 1      12.561.991: 1      14.116.397: 1      14.269.819: 1
## 1.286.863 : 1      1.333.227 : 1      15.857.031: 1      15.527.256: 1
## 13.617.993 : 1      1.693.224 : 1      1.637.681 : 1      17.131.713: 1
## 13.994.072 : 1      18.300.849: 1      2.054.026 : 1      1.826.027 : 1
## (Other)   :21      (Other)   :21      (Other)   :21      (Other)   :21
##          SMP2013          SMP2014          SMP2015
## 10.964.485: 1      11.657.610: 1      10.818.561: 1
## 12.441.544: 1      13.042.116: 1      11.859.347: 1
## 14.610.598: 1      15.350.853: 1      14.252.894: 1
## 16.214.100: 1      16.561.405: 1      14.774.824: 1
## 17.960.136: 1      18.719.145: 1      16.636.911: 1
## 1.928.978 : 1      2.020.278 : 1      1.840.885 : 1
## (Other)   :21      (Other)   :21      (Other)   :21
```

```
require(openxlsx)
```

```
## Loading required package: openxlsx
```

```
dadosJulio <- read.xlsx("Julio/INGESTÃO DE ALCOOL.xlsx")
head(dadosJulio)
```

```
##      data mes_ano dia_semana      horario uf  br    km
## 1 42736 janeiro      domingo 0.0000000000 SC 280  77,3
## 2 42736 janeiro      domingo 0.0006944444 PB 104   3,4
## 3 42736 janeiro      domingo 0.0208333333 TO 153 141,7
## 4 42736 janeiro      domingo 0.0902777778 SC 101   45
## 5 42736 janeiro      domingo 0.0208333333 GO  20  21,5
## 6 42736 janeiro      domingo 0.0972222222 SP 153  63,3
##
##                                     municipio
## 1 CORUPA
## 2 NOVA FLORESTA
## 3 ARAGUAINA
## 4 JOINVILLE
## 5 FORMOSA
## 6 SAO JOSE DO RIO PRETO
##
## 1 Saída de leito carroçável
## 2 Tombamento
## 3 Colisão traseira
## 4 Colisão traseira
## 5 Saída de leito carroçável
## 6 Colisão traseira
##      classificacao_acidente      fase_dia condicao_meteorologica
## 1      Com Vítimas Feridas Plena Noite      Chuva
## 2      Com Vítimas Fatais Plena Noite      Céu Claro
## 3      Com Vítimas Feridas Plena Noite      Ignorado
## 4      Com Vítimas Feridas Plena Noite      Chuva
## 5              Sem Vítimas Plena Noite      Céu Claro
## 6              Sem Vítimas Plena Noite      Céu Claro
##      tipo_pista      tracado_via      uso_solo
## 1 Simples      Não Informado      Não
## 2 Simples      Curva      Sim
## 3 Múltipla      Reta      Sim
## 4 Dupla      Curva      Sim
## 5 Simples      Interseção de vias      Não
## 6 Dupla      Curva      Sim
```

O excel guarda datas como número de dias a partir de Jan-0-1900. Vamos converter para data.

```
dadosJulio$data <- as.Date(dadosJulio$data , origin = '1899-12-30')
head(dadosJulio)
```

```
##      data mes_ano dia_semana      horario uf  br    km
## 1 2017-01-01 janeiro      domingo 0.0000000000 SC 280  77,3
## 2 2017-01-01 janeiro      domingo 0.0006944444 PB 104   3,4
## 3 2017-01-01 janeiro      domingo 0.0208333333 TO 153 141,7
## 4 2017-01-01 janeiro      domingo 0.0902777778 SC 101   45
## 5 2017-01-01 janeiro      domingo 0.0208333333 GO  20  21,5
## 6 2017-01-01 janeiro      domingo 0.0972222222 SP 153  63,3
##
##                                     municipio
## 1 CORUPA
## 2 NOVA FLORESTA
## 3 ARAGUAINA
## 4 JOINVILLE
## 5 FORMOSA
## 6 SAO JOSE DO RIO PRETO
```

```
##
## 1 Saída de leito carroçável
## 2 Tombamento
## 3 Colisão traseira
## 4 Colisão traseira
## 5 Saída de leito carroçável
## 6 Colisão traseira
## classificacao_acidente      fase_dia condicao_metereologica
## 1 Com Vítimas Feridas Plena Noite Chuva
## 2 Com Vítimas Fatais Plena Noite Céu Claro
## 3 Com Vítimas Feridas Plena Noite Ignorado
## 4 Com Vítimas Feridas Plena Noite Chuva
## 5 Sem Vítimas Plena Noite Céu Claro
## 6 Sem Vítimas Plena Noite Céu Claro
## tipo_pista      tracado_via uso_solo
## 1 Simples      Não Informado Não
## 2 Simples      Curva Sim
## 3 Múltipla      Reta Sim
## 4 Dupla      Curva Sim
## 5 Simples      Interseção de vias Não
## 6 Dupla      Curva Sim
```

Utilizando o argumento `sheet = 2`, deixamos indicado que iremos ler a tabela que está na segunda planilha do documento.

```
require(readxl)
```

```
## Loading required package: readxl
```

```
dadosMarony <- read_excel("Marony/EQ_10_anos_v2.xls",sheet = 2)
head(dadosMarony)
```

```
## # A tibble: 6 x 114
## SOLICITACAO CHAMADA PROGRAMA AREA_CONHECIMENTO COMITE NOME PROCESSO
## <chr> <chr> <chr> <chr> <chr> <chr> <chr>
## 1 4425040 PQ 2016 PROGRAMA ~ Tecnologia Química EQ Adela~ 303879/~
## 2 4420550 PQ 2016 PROGRAMA ~ Polímeros EQ Adria~ 301456/~
## 3 4426573 PQ 2016 PROGRAMA ~ Operações Caracte~ EQ Adria~ 305028/~
## 4 4429827 PQ 2016 PROGRAMA ~ Processos Bioquím~ EQ Ageno~ 307575/~
## 5 4420024 PQ 2016 PROGRAMA ~ Tecnologia Química EQ Alexs~ 301364/~
## 6 4424450 PQ 2016 PROGRAMA ~ Processos Bioquím~ EQ Aline~ 303462/~
## # ... with 107 more variables: MODALIDADE <chr>, INICIO <dtm>,
## # TERMINO <dtm>, TERMINO_PRIMEIRO_GD <chr>, DATA_PRIMEIRO_PQ <dtm>,
## # LIDER_GRUPO_PESQUISA <chr>, POSSUI_PDE_PREVIO <chr>,
## # INDICE_H_ISI_INFORMADO <dbl>, MEDIA_H_ISI_INF_NIVEL <dbl>,
## # INDICE_H_ISI_CALCULADO <dbl>, MEDIA_H_ISI_CAL_NIVEL <dbl>,
## # INDICE_H_SCOPUS_INFORMADO <lgl>, MEDIA_H_SCOPUS_INF_NIVEL <lgl>,
## # INDICE_H_SCOPUS_CALCULADO <dbl>, MEDIA_H_SCOPUS_CAL_NIVEL <dbl>,
## # CITACOES_ISI_INFORMADO <dbl>, MEDIA_CIT_ISI_INF_NIVEL <dbl>,
## # CITACOES_ISI_CALCULADO <dbl>, MEDIA_CIT_ISI_CAL_NIVEL <dbl>,
## # CITACOES_SCOPUS_INFORMADO <dbl>, MEDIA_CIT_SCOPUS_INF_NIVEL <dbl>,
## # CITACOES_SCOPUS_CALCULADO <dbl>, MEDIA_CIT_SCOPUS_CAL_NIVEL <dbl>,
## # PRODUCOES_EDUC_POP_CT <dbl>, MEMBRO_CORPO_EDITORIAL <dbl>,
## # ARTIGOS_TOTAL <dbl>, ART_JCR_MAIOR_IGUAL_6 <dbl>,
## # ART_JCR_5_E_MEIO_ATE_6 <dbl>, ART_JCR_5_ATE_5_E_MEIO <dbl>,
## # ART_JCR_4_E_MEIO_ATE_5 <dbl>, ART_JCR_4_ATE_4_E_MEIO <dbl>,
```

```
## # ART_JCR_3_E_MEIO_ATE_4 <dbl>, ART_JCR_3_ATE_3_E_MEIO <dbl>,
## # ART_JCR_2_E_MEIO_ATE_3 <dbl>, ART_JCR_2_ATE_2_E_MEIO <dbl>,
## # ART_JCR_1_E_MEIO_ATE_2 <dbl>, ART_JCR_1_ATE_1_E_MEIO <dbl>,
## # ART_JCR_MEIO_ATE_1 <dbl>, ART_JCR_ATE_MEIO <dbl>, ART_JCR_NULO <dbl>,
## # ART_QUALIS_A1 <dbl>, ART_QUALIS_A2 <dbl>, ART_QUALIS_B1 <dbl>,
## # ART_QUALIS_B2 <dbl>, ART_QUALIS_B3 <dbl>, ART_QUALIS_B4 <dbl>,
## # ART_QUALIS_B5 <dbl>, ART_QUALIS_C <dbl>, ART_QUALIS_NULO <dbl>,
## # ART_MENOR_IGUAL_10_AUTORES <dbl>, ART_ENTRE_11_19_AUTORES <dbl>,
## # ART_MAIOR_IGUAL_20_AUTORES <dbl>, ART_PRIMEIRO_AUTOR <dbl>,
## # ART_ULTIMO_AUTOR <dbl>, ART_JCR_1_3_PRIM_OU_ULT_AUT <dbl>,
## # ART_JCR_1_3_SEG_OU_PENULT_AUT <dbl>,
## # ART_JCR_MAIOR_3_PRI_OU_ULT_AUT <dbl>,
## # ART_JCR_MAIOR_3_SEG_OU_PEN_AUT <dbl>, MEDIA_ARTIGOS_PUB_NIVEL <dbl>,
## # SOMATORIO_GERAL_JCR <dbl>, MEDIA_GERAL_JCR <dbl>,
## # MEDIA_AUTORES_JCR_MAIOR_1 <dbl>, ORIENT_IC_CONCLUIDOS <dbl>,
## # ORIENT_GM_CONCLUIDOS <dbl>, ORIENT_GD_CONCLUIDOS <dbl>,
## # SUPERV_PD_CONCLUIDOS <dbl>, ORIENT_IC_ANDAMENTO <dbl>,
## # ORIENT_GM_ANDAMENTO <dbl>, ORIENT_GD_ANDAMENTO <dbl>,
## # SUPERV_PD_ANDAMENTO <dbl>, ORIENT_IC_CONCLUIDOS_TOT <dbl>,
## # ORIENT_GM_CONCLUIDOS_TOT <dbl>, ORIENT_GD_CONCLUIDOS_TOT <dbl>,
## # SUPERV_PD_CONCLUIDOS_TOT <dbl>, MEDIA_ORIENT_GD_NIVEL <dbl>,
## # MEDIA_ORIENT_GM_NIVEL <dbl>, TRAB_PUBLICADOS_EVENTOS <dbl>,
## # LIVROS <dbl>, CAPITULOS_LIVROS <dbl>, ORGANIZACAO_OBRA <dbl>,
## # PATENTES <dbl>, PATENTES_TOTAL <dbl>, EVENTOS_ORGANIZADOR <dbl>,
## # EVENTOS_PALESTRANTE_CONVIDADO <dbl>, PARTICIP_COMITE_ASSESS <dbl>,
## # INOVACOES_PROJETOS_EMPRESAS <dbl>, PREMIOS_E_TITULOS <dbl>,
## # PROJETOS_CERTIFICADOS <dbl>, DATA_INICIO_NIVEL_ATUAL <dtm>,
## # CURRICULO <chr>, NIVEL_ATUAL_PQ <chr>, CO_ORIENT_GM_CONCLUIDOS <dbl>,
## # CO_ORIENT_GD_CONCLUIDOS <dbl>, CO_ORIENT_GM_ANDAMENTO <dbl>,
## # CO_ORIENT_GD_ANDAMENTO <dbl>, CO_ORIENT_GM_CONCLUIDOS_TOT <dbl>,
## # CO_ORIENT_GD_CONCLUIDOS_TOT <dbl>, ORIENT_TCC_CONCLUIDAS <dbl>,
## # PARTICIP_BANCA_DOUT_CONC <dbl>, PARTICIP_BANCA_MESTR_CONC <dbl>, ...
```

```
dadosPedro <- read.csv("Pedro/saques_amostra_201701.csv",sep=";",encoding = "latin1")
head(dadosPedro)
```

```
##   UF codigo.siafi.municipio nome.municipio codigo.funcao codigo.subfuncao
## 1 SP           7107      SAO PAULO           8           244
## 2 SP           6313    CARAPICUIBA           8           244
## 3 BA           3541     IBICARAI            8           244
## 4 SP           7245    VOTUPORANGA           8           244
## 5 SP           6563  ITAQUAQUECETUBA           8           244
## 6 PE           2497     OURICURI            8           244
##   codigo.programa codigo.acao nis.favorecido      nome.favorecido
## 1           1335      8442  16482884963  MARILENE SILVA ALMEIDA
## 2           1335      8442  20710655384  GILDENE RITA DA SILVA
## 3           1335      8442  12318885348  VALDOMIRO MARTINS LISBOA
## 4           1335      8442  16015724421  ANA CARLA ALVES DA SILVA
## 5           1335      8442  20435890969  NAJARA DE JESUS MACEDO
## 6           1335      8442  16588231335  PATRICIA PEREIRA DA SILVA
##   fonte...fincalidade  mês.referencia.parcela Valor.Parcela
## 1 CAIXA - Programa Bolsa Família           1      227.00
## 2 CAIXA - Programa Bolsa Família           1       46.00
## 3 CAIXA - Programa Bolsa Família           1       85.00
## 4 CAIXA - Programa Bolsa Família           1     242.00
```

```
## 5 CAIXA - Programa Bolsa Família          1      124.00
## 6 CAIXA - Programa Bolsa Família          1      163.00
##   mês.competencia data.saque
## 1      jan/17 20/01/2017
## 2      jan/17 23/01/2017
## 3      jan/17 27/01/2017
## 4      jan/17 18/01/2017
## 5      jan/17 30/01/2017
## 6      jan/17 24/01/2017
```

```
str(dadosPedro$Valor.Parcela)
```

```
## Factor w/ 356 levels "101.00","1,022.00",...: 106 266 344 119 13 50 57 122 56 50 ...
```

A coluna “Valor.Parcela” esta utilizando separador de virgula para milhares e de ponto para decimais. Portanto, nao foi lida corretamente. Vamos retirar as virgulas referente aos milhares, e transformar em um vetor numérico.

```
dadosPedro$Valor.Parcela <- as.numeric( gsub(",", "", dadosPedro$Valor.Parcela))
```

```
str(dadosPedro$Valor.Parcela)
```

```
## num [1:10000] 227 46 85 242 124 163 171 246 170 163 ...
```

```
require(openxlsx)
```

```
dadosFernando <- read.xlsx("fernando/Planilha PNAE AF 2016.xlsx",startRow = 9)
```

```
head(dadosFernando)
```

```
##   Código.IBGE UF      Entidade.Executora Valor.transferido.FNDE
## 1   1200013 AC    PREF MUN DE ACRELANDIA      308233.4
## 2   1200104 AC    PREF MUN DE BRASILEIA      401400.5
## 3   1200179 AC    PREF MUN DE CAPIXABA      227697.6
## 4   1200203 AC    PREF MUN DE CRUZEIRO DO SUL  1294604.4
## 5   1200252 AC    PREF MUN DE EPITACIOLANDIA  309235.7
## 6   1200302 AC    PREF MUN DE FEIJO        489769.4
##   Valor.aquisições.agricultura.familiar Percentual
## 1                                8607.0 0.02792364
## 2                                0.0 0.00000000
## 3                                0.0 0.00000000
## 4                               369890.2 0.28571676
## 5                               86740.4 0.28049931
## 6                               65907.0 0.13456742
```

```
summary(dadosFernando)
```

```
##   Código.IBGE      UF      Entidade.Executora
## Length:5360      Length:5360      Length:5360
## Class :character  Class :character  Class :character
## Mode  :character  Mode  :character  Mode  :character
##
##
##
##   Valor.transferido.FNDE Valor.aquisições.agricultura.familiar
## Min.   :    531      Min.   :    0
## 1st Qu.:   77953      1st Qu.:  11009
## Median :  188931      Median :   38267
## Mean   :  724379      Mean   :  160220
## 3rd Qu.:  479738      3rd Qu.:  118781
```

```
## Max.      :176957074      Max.      :48854458
## Percentual
## Min.      :0.0000
## 1st Qu.   :0.1044
## Median    :0.2771
## Mean      :0.2771
## 3rd Qu.   :0.3842
## Max.      :1.0000
```