

Gas Prices in Brazil

```
library('tidyr')
library('dplyr')
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```
library('readr')
```

```
gasPrices_raw <- read.table(file = '2004-2019.tsv', sep = '\t', header = TRUE)
print(str(gasPrices_raw))
```

```
## 'data.frame':   106823 obs. of  21 variables:
##  $ X                : int  0 1 2 3 4 5 6 7 8 9 ...
##  $ DATA.INICIAL    : Factor w/ 785 levels "2004-05-09","2004-05-16",...: 1 1 1 1 1 1
##  $ DATA.FINAL      : Factor w/ 785 levels "2004-05-15","2004-05-22",...: 1 1 1 1 1 1
##  $ REGIÃO           : Factor w/ 5 levels "CENTRO OESTE",...: 1 1 1 1 2 2 2 2 2 ...
##  $ ESTADO           : Factor w/ 27 levels "ACRE","ALAGOAS",...: 7 9 11 12 2 5 6 10 15
##  $ PRODUTO           : Factor w/ 6 levels "Á"LEO DIESEL",...: 3 3 3 3 3 3 3 3 3 ...
##  $ NÚMERO.DE.POSTOS.PESQUISADOS : int  127 387 192 162 103 408 278 105 125 423 ...
##  $ UNIDADE.DE.MEDIDA : Factor w/ 3 levels "R$/13Kg","R$/l",...: 2 2 2 2 2 2 2 2 2 ..
##  $ PREÇO.MÉDIO.REVENDA : num  1.29 1.16 1.39 1.26 1.18 ...
##  $ DESVIO.PADRÃO.REVENDA : num  0.016 0.114 0.097 0.07 0.078 0.132 0.218 0.158 0.13 0.141
##  $ PREÇO.MÉDIO.NÍMIO.REVENDA : num  1.19 0.89 1.18 1.09 1.05 0.999 1.03 1.35 1.1 0.989 ...
##  $ PREÇO.MÉDIO.XÍMIO.REVENDA : num  1.35 1.45 1.76 1.51 1.4 ...
##  $ MARGEM.MÉDIA.REVENDA : Factor w/ 11930 levels "-", "0.001", "0.022",...: 395 331 351 364
##  $ COEF.DE.VARIAÇÃO.REVENDA : num  0.012 0.098 0.07 0.055 0.066 0.095 0.15 0.097 0.101 0.115
##  $ PREÇO.MÉDIO.DISTRIBUIÇÃO : Factor w/ 15997 levels "-", "0.506", "0.547",...: 220 158 365 225
##  $ DESVIO.PADRÃO.DISTRIBUIÇÃO : Factor w/ 5858 levels "-", "0", "0.001",...: 112 90 97 121 79 130
##  $ PREÇO.MÉDIO.NÍMIO.DISTRIBUIÇÃO : Factor w/ 21620 levels "-", "0.3257", "0.326",...: 20 115 194 250
##  $ PREÇO.MÉDIO.XÍMIO.DISTRIBUIÇÃO : Factor w/ 22576 levels "-", "0.5992", "0.7044",...: 221 381 596 7
##  $ COEF.DE.VARIAÇÃO.DISTRIBUIÇÃO : Factor w/ 397 levels "-", "0", "0.001",...: 135 117 100 145 84 13
##  $ MÊS              : int  5 5 5 5 5 5 5 5 5 5 ...
##  $ ANO               : int  2004 2004 2004 2004 2004 2004 2004 2004 2004 2004 ...
##  NULL
```

```
cat(c('\n\nColunas: ', names(gasPrices_raw)))
```

```
##
##
```

```
## Colunas:  X DATA.INICIAL DATA.FINAL REGIÃO ESTADO PRODUTO NÚMERO.DE.POSTOS.PESQUISADOS UNIDADE.DE.
```

```
print.data.frame(head(gasPrices_raw))
```

```
##   X DATA.INICIAL DATA.FINAL   REGIÃO          ESTADO
## 1 0   2004-05-09 2004-05-15 CENTRO OESTE  DISTRITO FEDERAL
## 2 1   2004-05-09 2004-05-15 CENTRO OESTE          GOIAS
```

```

## 3 2    2004-05-09 2004-05-15 CENTRO OESTE          MATO GROSSO
## 4 3    2004-05-09 2004-05-15 CENTRO OESTE MATO GROSSO DO SUL
## 5 4    2004-05-09 2004-05-15      NORDESTE          ALAGOAS
## 6 5    2004-05-09 2004-05-15      NORDESTE          BAHIA
##          PRODUTO NÃšMERO.DE.POSTOS.PESQUISADOS UNIDADE.DE.MEDIDA
## 1 ETANOL HIDRATADO                127                R$/1
## 2 ETANOL HIDRATADO                387                R$/1
## 3 ETANOL HIDRATADO                192                R$/1
## 4 ETANOL HIDRATADO                162                R$/1
## 5 ETANOL HIDRATADO                103                R$/1
## 6 ETANOL HIDRATADO                408                R$/1
##  PREÃ.O.MÃ.DIO.REVENDA DESVIO.PADRÃŋO.REVENDA PREÃ.O.MÃ.NIMO.REVENDA
## 1          1.288                0.016                1.190
## 2          1.162                0.114                0.890
## 3          1.389                0.097                1.180
## 4          1.262                0.070                1.090
## 5          1.181                0.078                1.050
## 6          1.383                0.132                0.999
##  PREÃ.O.MÃ.XIMO.REVENDA MARGEM.MÃ.DIA.REVENDA COEF.DE.VARIAÃ.ÃŋO.REVENDA
## 1          1.350                0.463                0.012
## 2          1.449                0.399                0.098
## 3          1.760                0.419                0.070
## 4          1.509                0.432                0.055
## 5          1.400                0.24                0.066
## 6          2.050                0.426                0.095
##  PREÃ.O.MÃ.DIO.DISTRIBUIÃ.ÃŋO DESVIO.PADRÃŋO.DISTRIBUIÃ.ÃŋO
## 1          0.825                0.11
## 2          0.763                0.088
## 3          0.97                0.095
## 4          0.83                0.119
## 5          0.941                0.077
## 6          0.957                0.128
##  PREÃ.O.MÃ.NIMO.DISTRIBUIÃ.ÃŋO PREÃ.O.MÃ.XIMO.DISTRIBUIÃ.ÃŋO
## 1          0.4201                0.9666
## 2          0.5013                1.05
## 3          0.5614                1.161
## 4          0.5991                1.22242
## 5          0.7441                1.0317
## 6          0.5686                1.35
##  COEF.DE.VARIAÃ.ÃŋO.DISTRIBUIÃ.ÃŋO MÃŠS ANO
## 1          0.133      5 2004
## 2          0.115      5 2004
## 3          0.098      5 2004
## 4          0.143      5 2004
## 5          0.082      5 2004
## 6          0.134      5 2004

```

```
gasPrices_temp <- gasPrices_raw
```

```

names(gasPrices_temp)[3] <- 'dataFinal'
names(gasPrices_temp)[4] <- 'regiao'
names(gasPrices_temp)[5] <- 'estado'
names(gasPrices_temp)[6] <- 'produto'
names(gasPrices_temp)[9] <- 'precoMedioRevenda'

```

```

levels(gasPrices_temp$produto)[1] <- 'OLEO DIESEL'
levels(gasPrices_temp$produto)[2] <- 'OLEO DIESEL S10'

gasPrices_temp$dataFinal <- as.Date(gasPrices_temp$dataFinal)

inicioAnalise <- as.Date('01/01/2016', '%d/%m/%Y')
produtoAnalizado <- 'GASOLINA COMUM'

gasPrices_allProducts <- gasPrices_temp %>%
  filter(dataFinal>inicioAnalise) %>%
  select(dataFinal, regiao, estado, produto, precoMedioRevenda) %>%
  arrange(estado) %>%
  arrange(regiao)

gasPrices <- gasPrices_allProducts %>%
  filter(produto==produtoAnalizado)

head(gasPrices)

```

```

##      dataFinal      regiao      estado      produto
## 1 2016-01-02 CENTRO OESTE DISTRITO FEDERAL GASOLINA COMUM
## 2 2016-01-09 CENTRO OESTE DISTRITO FEDERAL GASOLINA COMUM
## 3 2016-01-16 CENTRO OESTE DISTRITO FEDERAL GASOLINA COMUM
## 4 2016-01-23 CENTRO OESTE DISTRITO FEDERAL GASOLINA COMUM
## 5 2016-01-30 CENTRO OESTE DISTRITO FEDERAL GASOLINA COMUM
## 6 2016-02-06 CENTRO OESTE DISTRITO FEDERAL GASOLINA COMUM
##      precoMedioRevenda
## 1                3.770
## 2                3.772
## 3                3.967
## 4                3.968
## 5                3.967
## 6                3.967

```

```

for (regiao_str in unique(gasPrices$regiao)){
  anova <- aov(precoMedioRevenda ~ estado, gasPrices %>% filter(regiao==regiao_str))
  print('*****')
  print(regiao_str)
  print('*****')
  print(summary(anova))
  print(TukeyHSD(anova, ordered=TRUE))
  cat(rep('\n', 2))
}

```

```

## [1] "*****"
## [1] "CENTRO OESTE"
## [1] "*****"
##           Df Sum Sq Mean Sq F value    Pr(>F)
## estado      3   9.36   3.1183    20.8 6.16e-13 ***
## Residuals 728 109.15   0.1499
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Tukey multiple comparisons of means
## 95% family-wise confidence level

```

```

##      factor levels have been ordered
##
## Fit: aov(formula = precoMedioRevenda ~ estado, data = gasPrices %>% filter(regiao == regiao_str))
##
## $estado
##              diff              lwr              upr
## DISTRITO FEDERAL-MATO GROSSO DO SUL 0.14843716  0.04420348 0.2526708
## MATO GROSSO-MATO GROSSO DO SUL      0.21748634  0.11325266 0.3217200
## GOIAS-MATO GROSSO DO SUL             0.30966120  0.20542752 0.4138949
## MATO GROSSO-DISTRITO FEDERAL         0.06904918 -0.03518450 0.1732829
## GOIAS-DISTRITO FEDERAL               0.16122404  0.05699037 0.2654577
## GOIAS-MATO GROSSO                   0.09217486 -0.01205881 0.1964085
##              p adj
## DISTRITO FEDERAL-MATO GROSSO DO SUL 0.0014990
## MATO GROSSO-MATO GROSSO DO SUL      0.0000006
## GOIAS-MATO GROSSO DO SUL             0.0000000
## MATO GROSSO-DISTRITO FEDERAL         0.3213437
## GOIAS-DISTRITO FEDERAL               0.0004347
## GOIAS-MATO GROSSO                   0.1043393
##
##
## [1] "*****"
## [1] "NORDESTE"
## [1] "*****"
##              Df Sum Sq Mean Sq F value Pr(>F)
## estado          8  17.82   2.2270   16.17 <2e-16 ***
## Residuals     1638  225.63   0.1377
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##      Tukey multiple comparisons of means
##      95% family-wise confidence level
##      factor levels have been ordered
##
## Fit: aov(formula = precoMedioRevenda ~ estado, data = gasPrices %>% filter(regiao == regiao_str))
##
## $estado
##              diff              lwr              upr              p adj
## PARAIBA-MARANHAO      0.048267760 -0.07224016 0.1687757 0.9465558
## PERNAMBUCO-MARANHAO   0.102502732 -0.01800518 0.2230106 0.1702072
## PIAUI-MARANHAO        0.147617486  0.02710957 0.2681254 0.0046438
## SERGIPE-MARANHAO      0.152426230  0.03191831 0.2729341 0.0028672
## RIO GRANDE DO NORTE-MARANHAO 0.242038251  0.12153034 0.3625462 0.0000000
## BAHIA-MARANHAO        0.255890710  0.13538279 0.3763986 0.0000000
## ALAGOAS-MARANHAO      0.283683060  0.16317514 0.4041910 0.0000000
## CEARA-MARANHAO        0.321142077  0.20063416 0.4416500 0.0000000
## PERNAMBUCO-PARAIBA    0.054234973 -0.06627294 0.1747429 0.8988894
## PIAUI-PARAIBA         0.099349727 -0.02115819 0.2198576 0.2042875
## SERGIPE-PARAIBA       0.104158470 -0.01634945 0.2246664 0.1540422
## RIO GRANDE DO NORTE-PARAIBA 0.193770492  0.07326258 0.3142784 0.0000231
## BAHIA-PARAIBA         0.207622951  0.08711503 0.3281309 0.0000036
## ALAGOAS-PARAIBA       0.235415301  0.11490738 0.3559232 0.0000001
## CEARA-PARAIBA         0.272874317  0.15236640 0.3933822 0.0000000
## PIAUI-PERNAMBUCO      0.045114754 -0.07539316 0.1656227 0.9640104

```

```

## SERGIPE-PERNAMBUCO      0.049923497 -0.07058442 0.1704314 0.9353339
## RIO GRANDE DO NORTE-PERNAMBUCO 0.139535519 0.01902760 0.2600434 0.0100359
## BAHIA-PERNAMBUCO        0.153387978 0.03288006 0.2738959 0.0025982
## ALAGOAS-PERNAMBUCO      0.181180328 0.06067241 0.3016882 0.0001133
## CEARA-PERNAMBUCO        0.218639344 0.09813143 0.3391473 0.0000007
## SERGIPE-PIAUI           0.004808743 -0.11569917 0.1253167 1.0000000
## RIO GRANDE DO NORTE-PIAUI 0.094420765 -0.02608715 0.2149287 0.2663271
## BAHIA-PIAUI             0.108273224 -0.01223469 0.2287811 0.1188269
## ALAGOAS-PIAUI           0.136065574 0.01555766 0.2565735 0.0137535
## CEARA-PIAUI             0.173524590 0.05301667 0.2940325 0.0002834
## RIO GRANDE DO NORTE-SERGIPE 0.089612022 -0.03089589 0.2101199 0.3367303
## BAHIA-SERGIPE           0.103464481 -0.01704344 0.2239724 0.1606746
## ALAGOAS-SERGIPE         0.131256831 0.01074891 0.2517647 0.0209453
## CEARA-SERGIPE           0.168715847 0.04820793 0.2892238 0.0004937
## BAHIA-RIO GRANDE DO NORTE 0.013852459 -0.10665546 0.1343604 0.9999926
## ALAGOAS-RIO GRANDE DO NORTE 0.041644809 -0.07886311 0.1621527 0.9779497
## CEARA-RIO GRANDE DO NORTE 0.079103825 -0.04140409 0.1996117 0.5162561
## ALAGOAS-BAHIA           0.027792350 -0.09271557 0.1483003 0.9985773
## CEARA-BAHIA             0.065251366 -0.05525655 0.1857593 0.7578519
## CEARA-ALAGOAS           0.037459016 -0.08304890 0.1579669 0.9888248
##
##
## [1] "*****"
## [1] "NORTE"
## [1] "*****"
##           Df Sum Sq Mean Sq F value Pr(>F)
## estado      6  53.16   8.860    77.25 <2e-16 ***
## Residuals 1274 146.12   0.115
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##     factor levels have been ordered
##
## Fit: aov(formula = precoMedioRevenda ~ estado, data = gasPrices %>% filter(regiao == regiao_str))
##
## $estado
##           diff           lwr           upr           p adj
## RORAIMA-AMAPA 0.15127322 0.04672182 0.2558246 0.0004145
## Amazonas-AMAPA 0.18975410 0.08520269 0.2943055 0.0000021
## RONDONIA-AMAPA 0.32882514 0.22427373 0.4333765 0.0000000
## TOCANTINS-AMAPA 0.35806557 0.25351417 0.4626170 0.0000000
## PARA-AMAPA    0.37383607 0.26928466 0.4783875 0.0000000
## ACRE-AMAPA    0.69568852 0.59113712 0.8002399 0.0000000
## Amazonas-RORAIMA 0.03848087 -0.06607053 0.1430323 0.9320429
## RONDONIA-RORAIMA 0.17755191 0.07300051 0.2821033 0.0000125
## TOCANTINS-RORAIMA 0.20679235 0.10224094 0.3113438 0.0000001
## PARA-RORAIMA  0.22256284 0.11801143 0.3271142 0.0000000
## ACRE-RORAIMA  0.54441530 0.43986389 0.6489667 0.0000000
## RONDONIA-AMAZONAS 0.13907104 0.03451963 0.2436224 0.0017438
## TOCANTINS-AMAZONAS 0.16831148 0.06376007 0.2728629 0.0000456
## PARA-AMAZONAS 0.18408197 0.07953056 0.2886334 0.0000048
## ACRE-AMAZONAS 0.50593443 0.40138302 0.6104858 0.0000000

```

```

## TOCANTINS-RONDONIA 0.02924044 -0.07531097 0.1337918 0.9823062
## PARA-RONDONIA      0.04501093 -0.05954048 0.1495623 0.8650303
## ACRE-RONDONIA      0.36686339  0.26231198 0.4714148 0.0000000
## PARA-TOCANTINS     0.01577049 -0.08878092 0.1203219 0.9994104
## ACRE-TOCANTINS     0.33762295  0.23307154 0.4421744 0.0000000
## ACRE-PARA          0.32185246  0.21730105 0.4264039 0.0000000
##
##
## [1] "*****"
## [1] "SUDESTE"
## [1] "*****"
##           Df Sum Sq Mean Sq F value Pr(>F)
## estado      3  30.81   10.27   60.39 <2e-16 ***
## Residuals   728 123.82    0.17
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Tukey multiple comparisons of means
## 95% family-wise confidence level
## factor levels have been ordered
##
## Fit: aov(formula = precoMedioRevenda ~ estado, data = gasPrices %>% filter(regiao == regiao_str))
##
## $estado
##           diff          lwr          upr          p adj
## ESPIRITO SANTO-SAO PAULO 0.2037705 0.09275247 0.3147885 0.0000163
## MINAS GERAIS-SAO PAULO  0.3489126 0.23789455 0.4599306 0.0000000
## RIO DE JANEIRO-SAO PAULO 0.5618306 0.45081258 0.6728486 0.0000000
## MINAS GERAIS-ESPIRITO SANTO 0.1451421 0.03412405 0.2561601 0.0044456
## RIO DE JANEIRO-ESPIRITO SANTO 0.3580601 0.24704209 0.4690781 0.0000000
## RIO DE JANEIRO-MINAS GERAIS 0.2129180 0.10190001 0.3239361 0.0000058
##
##
## [1] "*****"
## [1] "SUL"
## [1] "*****"
##           Df Sum Sq Mean Sq F value Pr(>F)
## estado      2  13.48    6.740   59.97 <2e-16 ***
## Residuals   546  61.36    0.112
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Tukey multiple comparisons of means
## 95% family-wise confidence level
## factor levels have been ordered
##
## Fit: aov(formula = precoMedioRevenda ~ estado, data = gasPrices %>% filter(regiao == regiao_str))
##
## $estado
##           diff          lwr          upr          p adj
## PARANA-SANTA CATARINA 0.09860656 0.01624198 0.1809711 0.0140376
## RIO GRANDE DO SUL-SANTA CATARINA 0.37055738 0.28819280 0.4529220 0.0000000
## RIO GRANDE DO SUL-PARANA 0.27195082 0.18958624 0.3543154 0.0000000
##

```

```
##
##
meanByState <- gasPrices %>%
  group_by(regiao, estado) %>%
  summarize(mean = mean(precoMedioRevenda)) %>%
  top_n(2, mean)

print.data.frame(meanByState)
```

```
##      regiao      estado      mean
## 1 CENTRO OESTE      GOIAS 4.167951
## 2 CENTRO OESTE      MATO GROSSO 4.075776
## 3 NORDESTE      ALAGOAS 4.130492
## 4 NORDESTE      CEARA 4.167951
## 5 NORTE      ACRE 4.529377
## 6 NORTE      PARA 4.207525
## 7 SUDESTE      MINAS GERAIS 4.156383
## 8 SUDESTE      RIO DE JANEIRO 4.369301
## 9 SUL      PARANA 3.908246
## 10 SUL RIO GRANDE DO SUL 4.180197
```

```
for (prod in unique(gasPrices_allProducts$produto)){
topMeanValues <- gasPrices_allProducts %>%
  filter(produto==prod) %>%
  select(estado, precoMedioRevenda) %>%
  group_by(estado) %>%
  summarize(mean = mean(precoMedioRevenda)) %>%
  arrange(desc(mean))
cat(c('Produto: ', prod, '\n'))
print.data.frame(topMeanValues[1:5,])
cat('\n')
}
```

```
## Produto: ETANOL HIDRATADO
##      estado      mean
## 1 RIO GRANDE DO SUL 3.767492
## 2      AMAPA 3.743867
## 3 RORAIMA 3.714339
## 4      ACRE 3.690464
## 5      PARA 3.628126
##
```

```
## Produto: OLEO DIESEL
##      estado      mean
## 1      ACRE 4.022169
## 2      AMAPA 3.729202
## 3 MATO GROSSO 3.565060
## 4      PARA 3.514683
## 5 RONDONIA 3.489650
##
```

```
## Produto: OLEO DIESEL S10
##      estado      mean
## 1      AMAPA 4.210208
## 2      ACRE 4.090852
## 3 MATO GROSSO 3.677940
```

```

## 4 MATO GROSSO DO SUL 3.607344
## 5          RONDONIA 3.583765
##
## Produto:  GASOLINA COMUM
##          estado      mean
## 1          ACRE 4.529377
## 2    RIO DE JANEIRO 4.369301
## 3          PARA 4.207525
## 4          TOCANTINS 4.191754
## 5 RIO GRANDE DO SUL 4.180197
##
## Produto:  GLP
##          estado      mean
## 1 MATO GROSSO 86.13658
## 2    TOCANTINS 75.60252
## 3          AMAPA 71.56960
## 4          RORAIMA 71.42717
## 5          ACRE 69.77173
##
## Produto:  GNV
##          estado      mean
## 1          TOCANTINS 3.381750
## 2          PIAUI 3.311200
## 3          MARANHAO 3.300333
## 4  DISTRITO FEDERAL 3.279250
## 5          AMAZONAS 2.974810

```