

Trabalho Final

Eduardo Santiago

17/12/2019

Análise dos Suicídios no Brasil

Objetivo

Essa análise tem como principal objetivo encontrar os fatores de risco para suicídio no Brasil e prever a evolução desses dados para os próximos anos.

A Base de Dados

Nessa análise será usada a base de dados de mortalidade da **OMS** Mortality, ICD-10. Ela é organizada pelo código dos países como variáveis categóricas e está em formato **long**. Com os dados dos países, outras variáveis categóricas são o tipo de morte e o gênero. Com isso, o número de mortes é separado por faixas-etárias padronizadas pela **OMS**.

Fonte dos Dados

Os dados foram obtidos do site internacional da OMS https://www.who.int/healthinfo/statistics/mortality_rawdata/en/ Onde eles podem ser baixado em formato “puro”, como dois arquivos com mais de 10Mb para serem tratados. Eles vem como arquivos CSV que, após terem as causas da morte e países desejados selecionados, serão unidos e transformados em um único arquivo. Agilizando o tratamento dos dados.

Dicionário de Dados

O dicionário de dados também está no link https://www.who.int/healthinfo/statistics/mortality_rawdata/en/ Apesar do dicionário conter as informações sobre todas as bases de dados, nessa sessão só serão exibidas as tabelas e linhas utilizadas.

Perguntas em Potencial

- Existe uma diferença significativa entre as taxas de suicídio entre os gêneros?
- Quais são as categorias com maior chance de cometerem suicídios?
- Utilizando um modelo de regressão linear, as categorias em mais risco estão com o número de suicídios aumentando ou diminuindo?

Análise Exploratória

Análise Exploratória dos Suicídios no Brasil

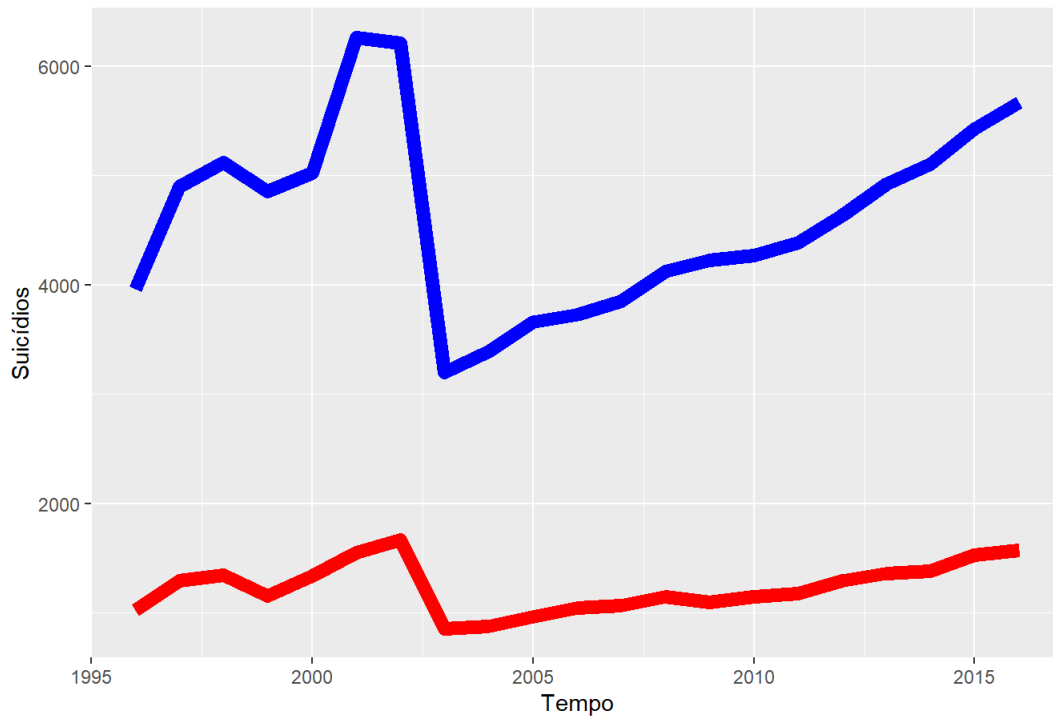
Este relatório irá abordar insights sobre o número de suicídios no Brasil, os métodos mais utilizados e que categorias estão em mais risco.

Perguntas:

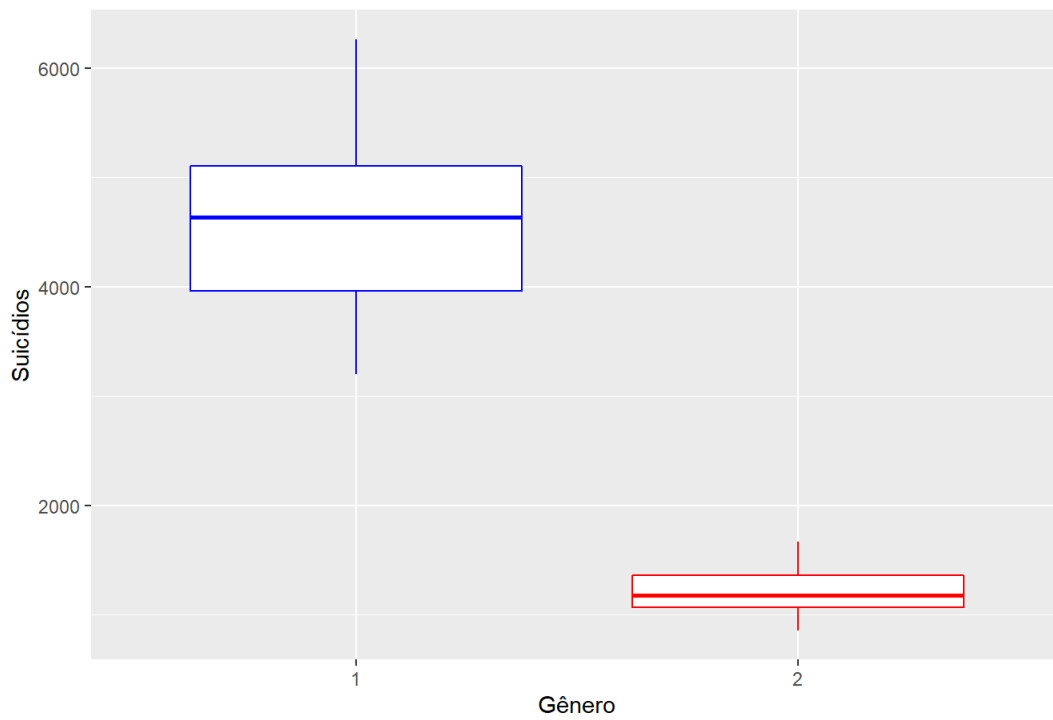
- 1 - Quem cometeu mais suicídios? Existe uma diferença significativa?
- 2 - Existe uma preferência por métodos de suicídio por gênero?
- 3 - Qual a faixa etária mais suicida?

1 - Quem cometeu mais suicídios? Existe uma diferença significativa?

Suicídios dividos por gênero

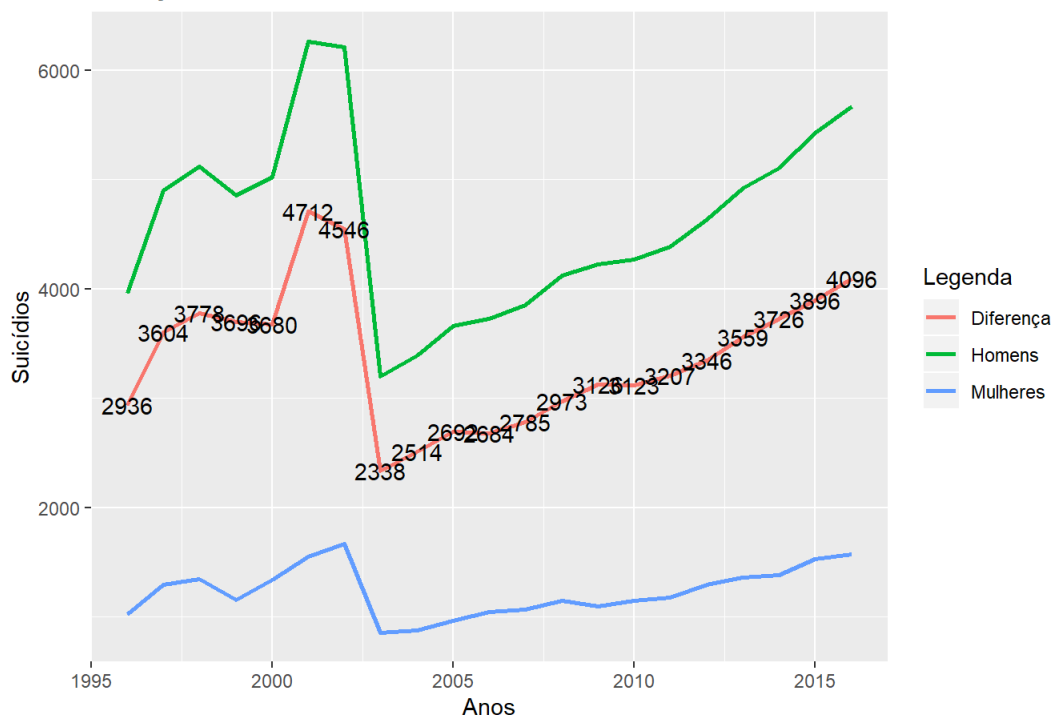


Suicídios dividos por gênero



É possível perceber uma diferença significativa no número de suicídios entre homens(1) e mulheres(2)

Diferença total dos suicídios

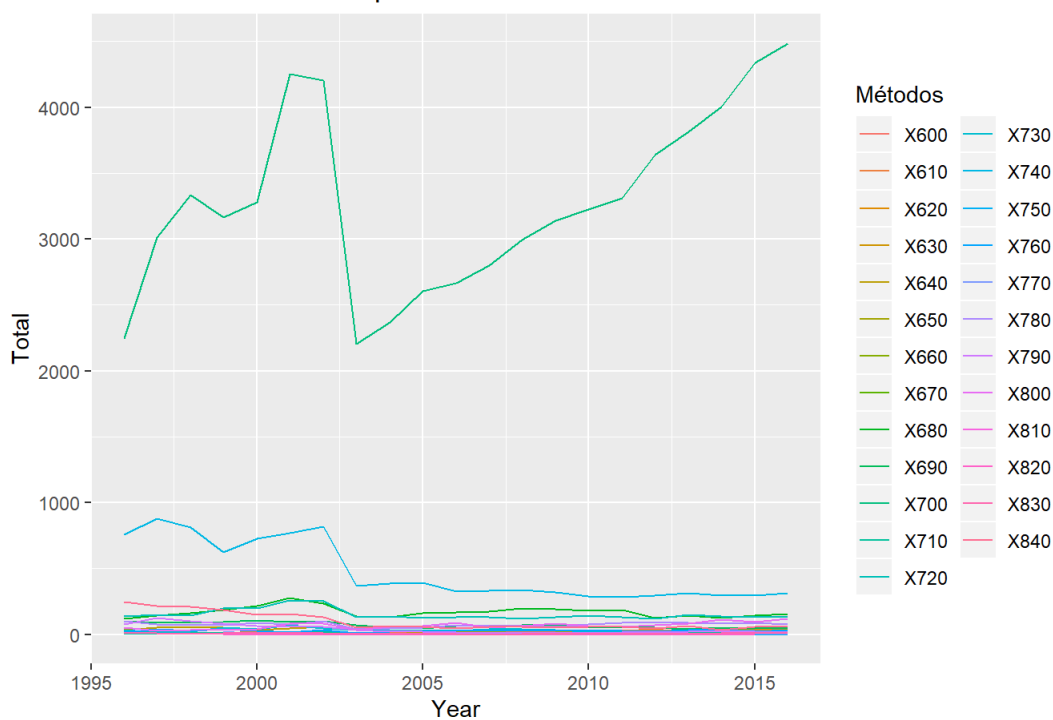


Neste gráfico temos a diferença “homens-mulheres” no mesmo gráfico, representada por uma linha entre os dois.

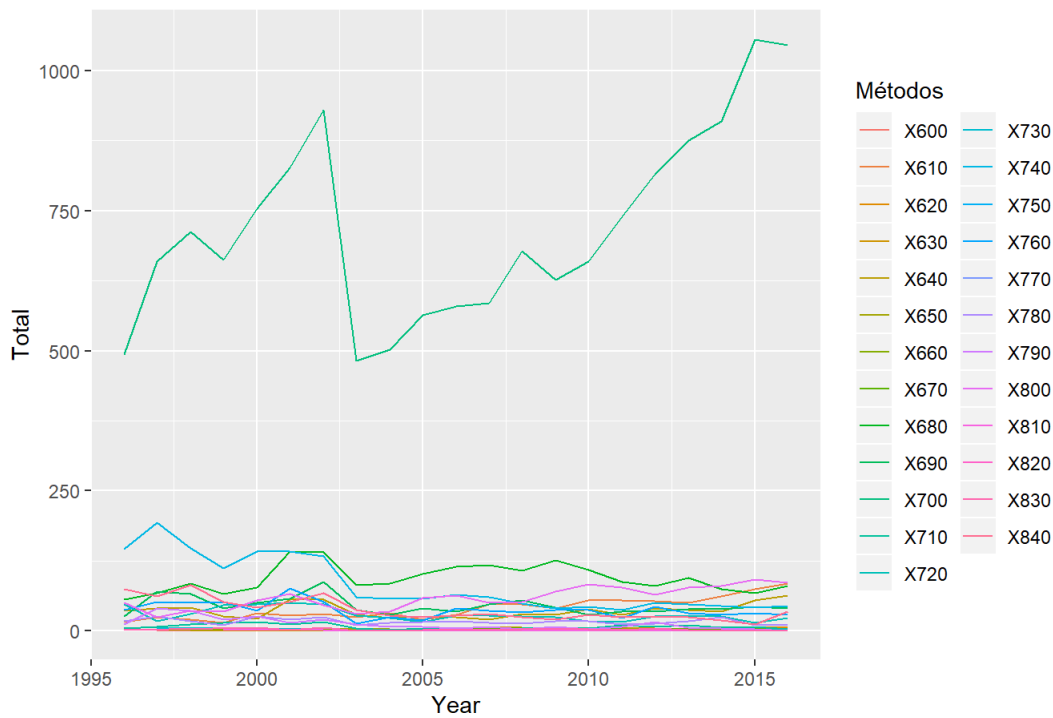
Assim sendo, é possível determinar uma diferença significativa entre a taxa de suicídio de homens e mulheres.

2 - Existe uma preferência por métodos de suicídio por gênero?

Métodos mais comuns para homens

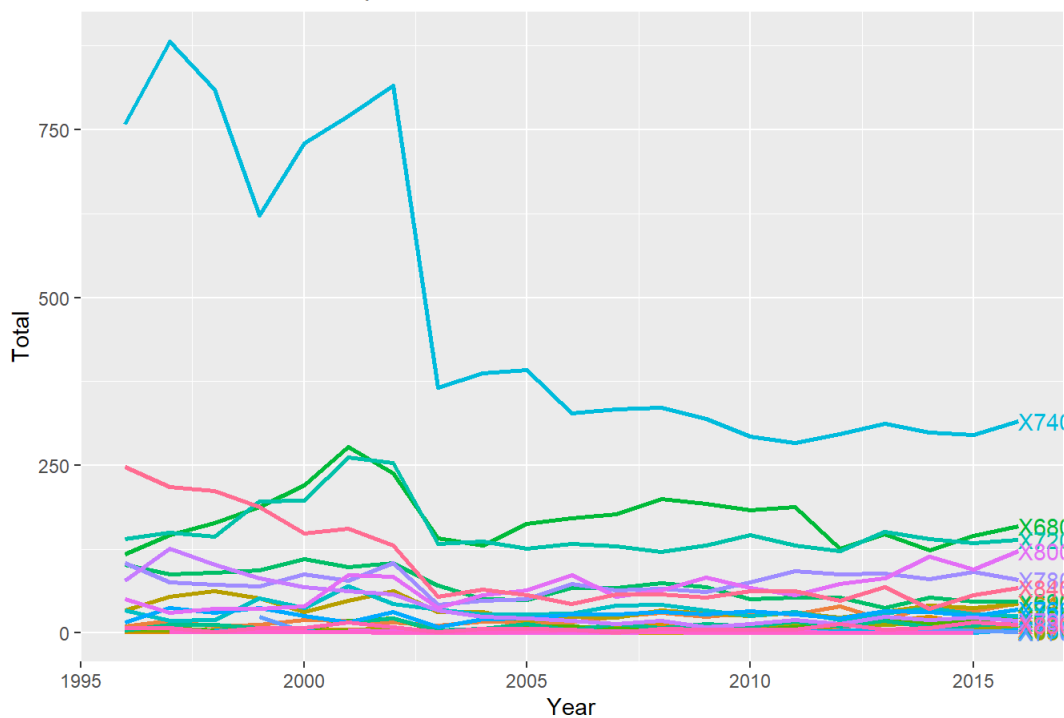


Métodos mais comuns para mulheres

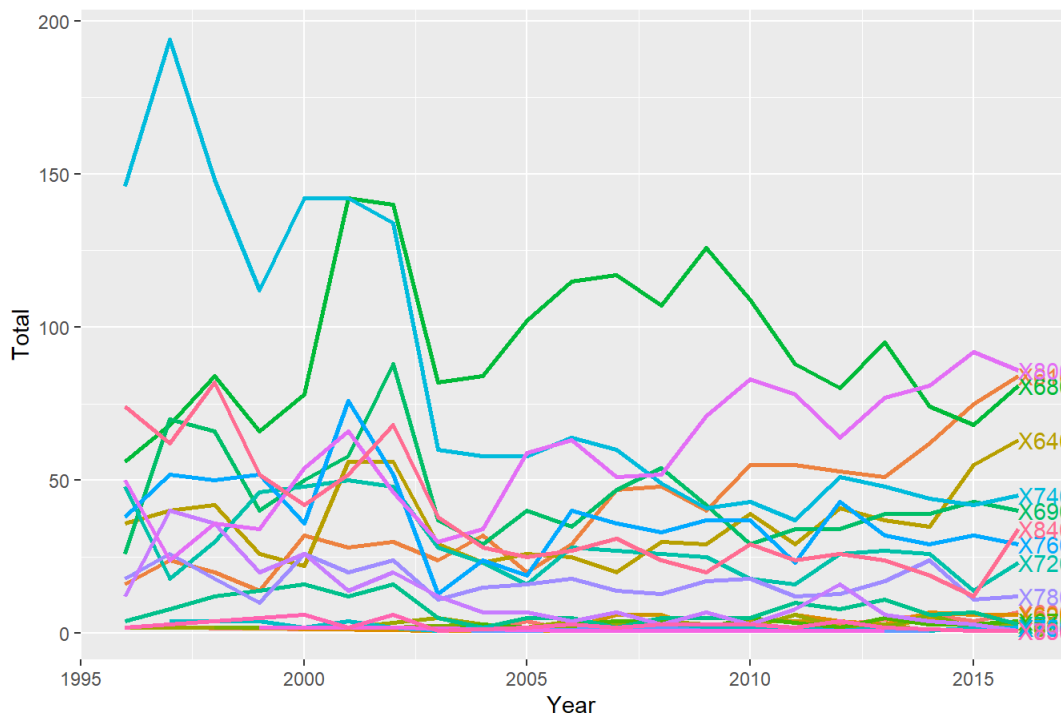


O método mais comum em questão é o mesmo para os dois gêneros, o enforcamento/estrangulamento. Essa forma de suicídio é relativamente comum pela sua facilidade de conseguir materiais e simples execução.

Métodos mais comuns para homens, excluindo enforcamento



Métodos mais comuns para mulheres, excluindo enforcamento

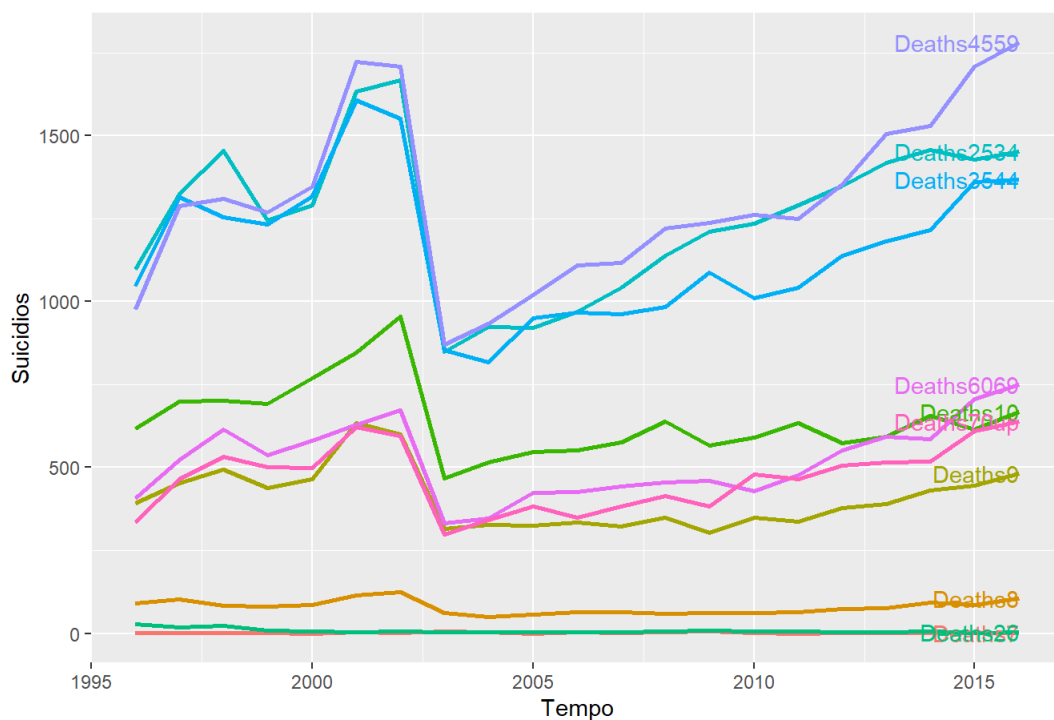


Agora temos diferenças significativas nas formas de suicídio separando por gênero, enquanto os homens optam por “X740”, representando armas de fogo, e bem abaixo temos envenenamento e quedas. As mulheres preferem se suicidar pulando de locais altos e com envenenamento. Essa diferença é um fato importante quando estamos tratando de número de suicídios por gênero, haja vista que armas de fogo apresentam uma chance de sucesso relativamente mais alta do que overdose por remédios, drogas, etc. Contribuindo para os maiores números de suicídios com sucesso pelo sexo masculino. Por outro lado, a quantidade bruta anual não é significativa para indicar que esse seja o único fator resultante nessa discrepância.

Uma observação importante foi a redução drástica no número de suicídios não só com armas de fogo, mas geral do ano de 2003 em diante. O ano da campanha do desarmamento teve um impacto positivo na prevenção ao suicídio no país.

3 - Qual a faixa etária mais suicida?

Faixas etárias em risco



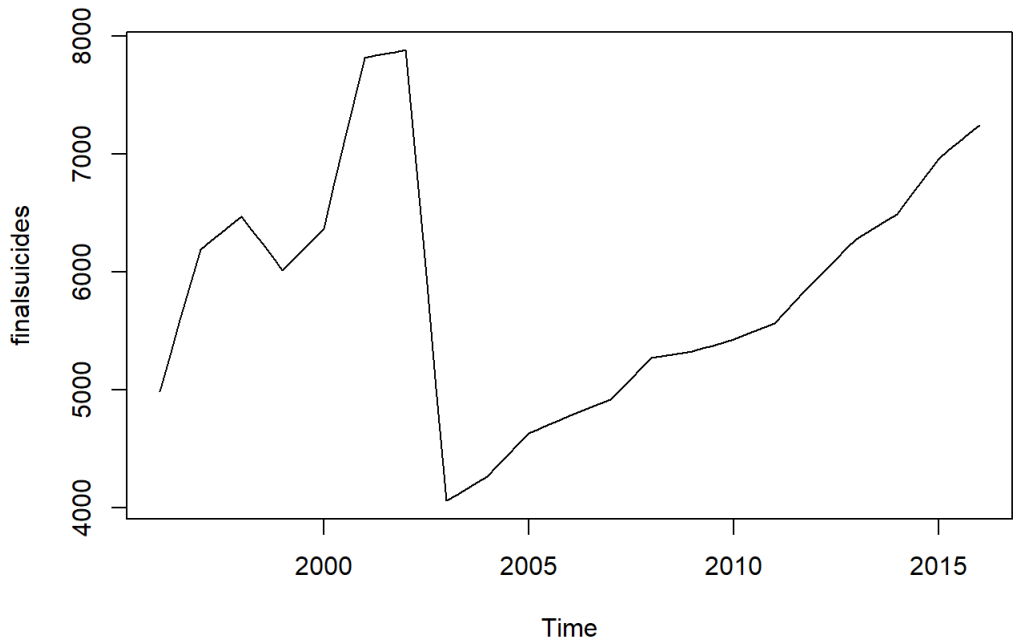
Diferentemente do senso comum, que dita adolescentes e idosos como faixas etárias de risco. o gráfico mostra como pessoas entre 25-59 anos representam significativamente taxas de suicídio elevadas. Essa parcela da sociedade tem mais responsabilidade e normalmente possuem uma rede de apoio reduzida. Fatores estes que contribuem para as taxas de suicídio.

Conclusão da Análise Exploratória

Com esse relatório é possível deduzir que homens de meia idade são as categorias com mais risco de suicídio. Seus métodos são mais letais do que os das mulheres e o número de pessoas que tiram as suas próprias vidas está aumentando cada vez mais, mesmo após a redução drástica da metade da década de 00.

Modelagem

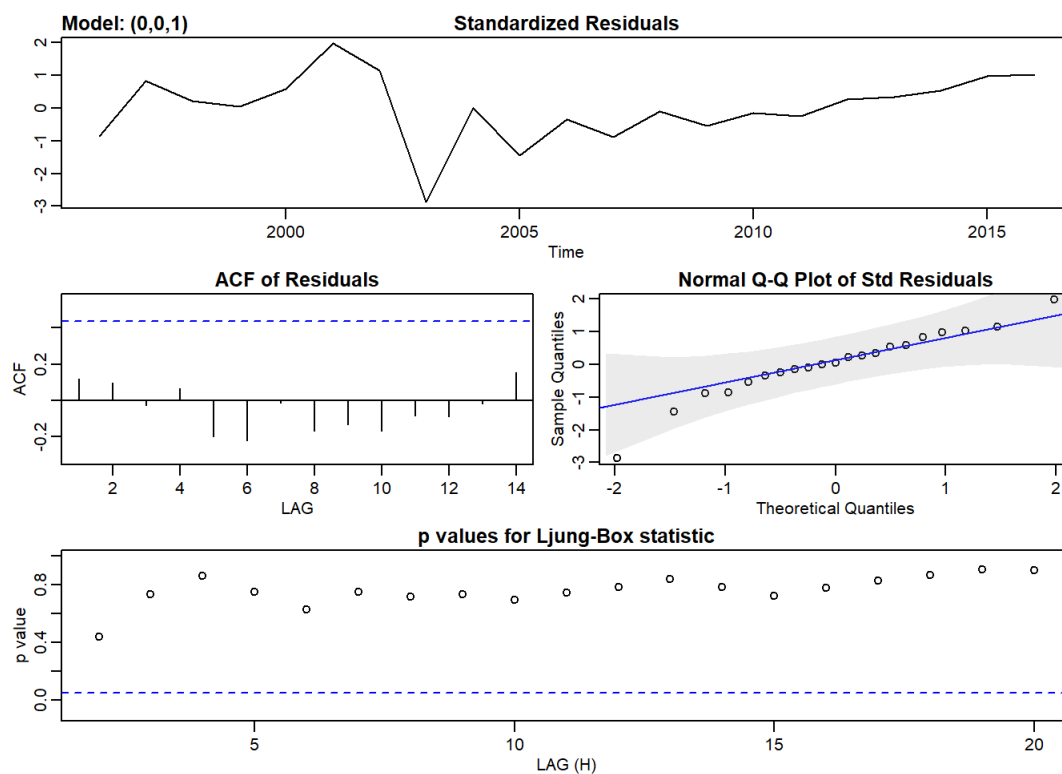
Os dados dos suicídios nos últimos anos podem ser vistos da seguinte forma



Analisando o gráfico, fica evidente a necessidade de um tratamento desses dados para que seja realizada uma previsão temporal. Dessa forma, vários modelos foram gerados e testados em níveis de diferenciação diferentes.

O BDD trouxe os dados divididos anualmente, dessa forma foi utilizado a interpolação para gerar novos dados trimestrais. Sendo mais fácil de visualizar as previsões realizadas.

```
## initial value 6.962701
## iter 2 value 6.764503
## iter 3 value 6.755053
## iter 4 value 6.751528
## iter 5 value 6.751036
## iter 6 value 6.750518
## iter 7 value 6.750456
## iter 8 value 6.750434
## iter 9 value 6.750423
## iter 10 value 6.750420
## iter 10 value 6.750420
## iter 10 value 6.750420
## final value 6.750420
## converged
## initial value 6.749239
## iter 2 value 6.748746
## iter 3 value 6.748644
## iter 4 value 6.748644
## iter 5 value 6.748638
## iter 5 value 6.748638
## iter 5 value 6.748638
## final value 6.748638
## converged
```

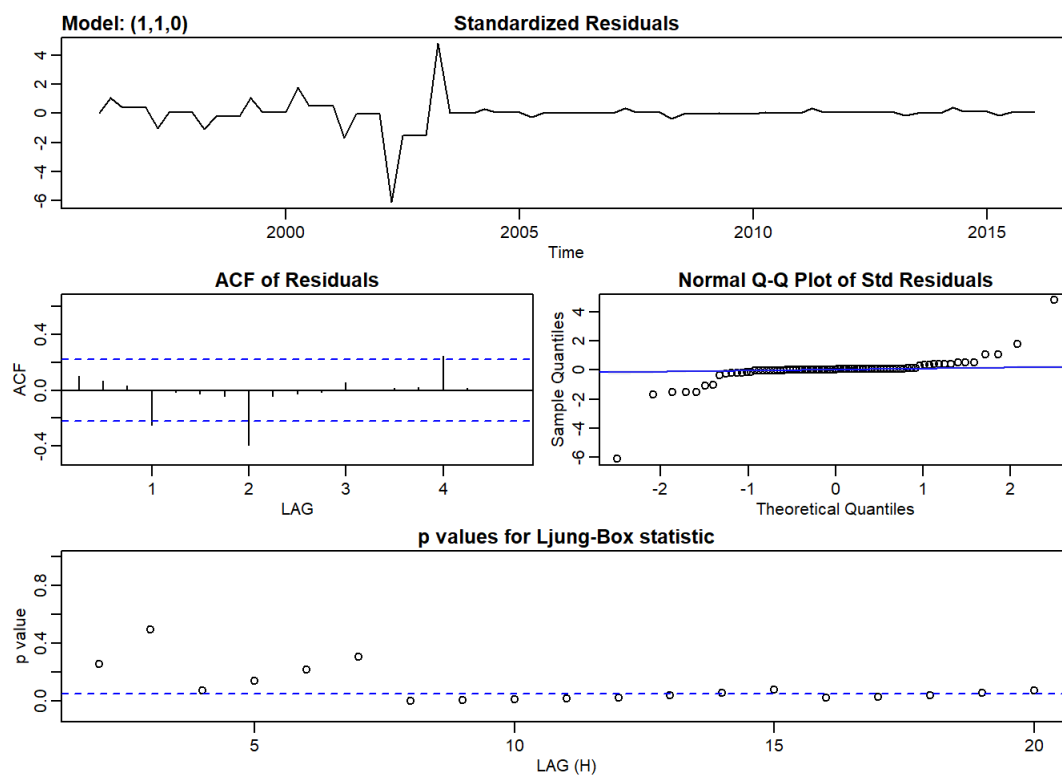


```
## $fit
##
## Call:
## stats::arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D,
##     Q), period = S), xreg = xmean, include.mean = FALSE, transform.pars = trans,
##     fixed = fixed, optim.control = list(trace = trc, REPORT = 1, reltol = tol))
##
## Coefficients:
##          mal          xmean
##      0.6513    5853.5354
## s.e.  0.1456    297.7915
##
## sigma^2 estimated as 708563:  log likelihood = -171.52,  aic = 349.04
##
## $degrees_of_freedom
## [1] 19
##
## $ttable
##      Estimate      SE t.value p.value
## mal      0.6513   0.1456  4.4727  3e-04
## xmean 5853.5354 297.7915 19.6565  0e+00
##
## $AIC
## [1] 16.62087
##
## $AICc
## [1] 16.65261
##
## $BIC
## [1] 16.77008
```

```

## initial value 5.502929
## iter 2 value 5.073440
## iter 3 value 5.073319
## iter 4 value 5.073244
## iter 5 value 5.073227
## iter 6 value 5.073204
## iter 7 value 5.073203
## iter 8 value 5.073202
## iter 9 value 5.073202
## iter 10 value 5.073202
## iter 10 value 5.073202
## iter 10 value 5.073202
## final value 5.073202
## converged
## initial value 5.080799
## iter 2 value 5.080542
## iter 3 value 5.080091
## iter 4 value 5.080085
## iter 5 value 5.080083
## iter 6 value 5.080083
## iter 6 value 5.080083
## final value 5.080083
## converged

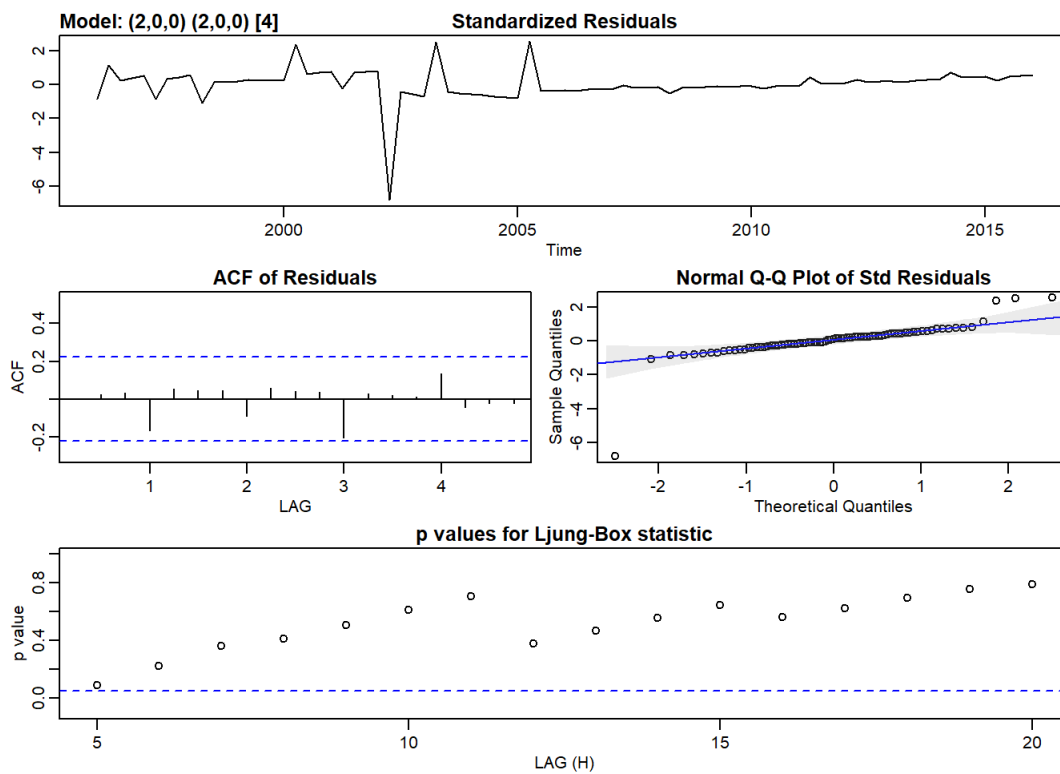
```




```
## $fit
##
## Call:
## stats::arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D,
##      Q), period = S), xreg = constant, transform.pars = trans, fixed = fixed,
##      optim.control = list(trace = trc, REPORT = 1, reltol = tol))
##
## Coefficients:
##          ar1  constant
##      0.7560   39.6663
## s.e.  0.0717   70.7220
##
## sigma^2 estimated as 25580:  log likelihood = -519.92,  aic = 1045.84
##
## $degrees_of_freedom
## [1] 78
##
## $ttable
##           Estimate      SE t.value p.value
## ar1           0.7560  0.0717 10.5490  0.0000
## constant    39.6663 70.7220  0.5609  0.5765
##
## $AIC
## [1] 13.07304
##
## $AICc
## [1] 13.07499
##
## $BIC
## [1] 13.16237
```

Os dois primeiros apresentaram altas taxas de AIC e BIC, apesar dos gráficos de diferenciação estarem intuitivamente mais corretos. Por outro lado, caso não houvesse nenhuma diferenciação, e a utilização de um modelo SAR(2). As taxas de AIC e BIC se mostraram comparativamente baixas.

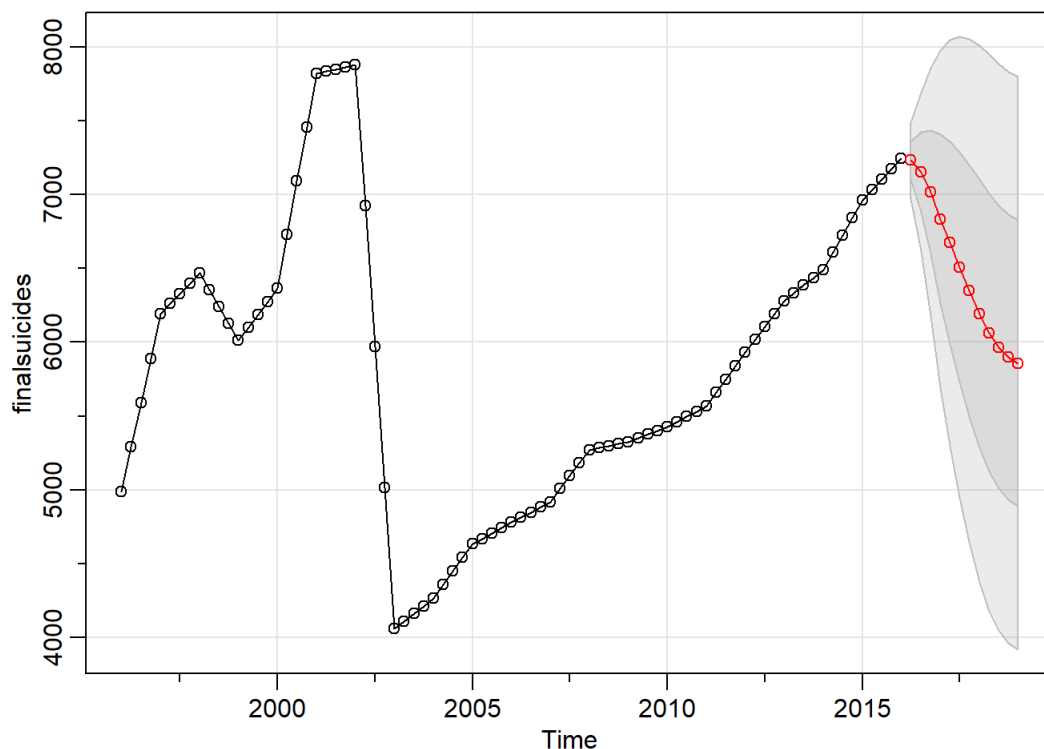
```
## initial  value 6.934196
## iter    2 value 6.792358
## iter    3 value 5.902010
## iter    4 value 5.836254
## iter    5 value 5.570306
## iter    6 value 5.444457
## iter    7 value 5.346332
## iter    8 value 5.191338
## iter    9 value 5.117222
## iter   10 value 4.928794
## iter   11 value 4.901858
## iter   12 value 4.896238
## iter   13 value 4.896186
## iter   14 value 4.894831
## iter   15 value 4.894141
## iter   16 value 4.894117
## iter   17 value 4.894049
## iter   18 value 4.894037
## iter   19 value 4.894036
## iter   19 value 4.894036
## iter   19 value 4.894036
## final   value 4.894036
## converged
## initial  value 4.906595
## iter    2 value 4.906300
## iter    3 value 4.905859
## iter    4 value 4.905704
## iter    5 value 4.905680
## iter    6 value 4.905674
## iter    7 value 4.905668
## iter    8 value 4.905666
## iter    9 value 4.905666
## iter    9 value 4.905666
## iter    9 value 4.905666
## final   value 4.905666
## converged
```



```
## $fit
##
## Call:
## stats::arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D,
##     Q), period = S), xreg = xmean, include.mean = FALSE, transform.pars = trans,
##     fixed = fixed, optim.control = list(trace = trc, REPORT = 1, reltol = tol))
##
## Coefficients:
##          ar1          ar2          sar1          sar2          xmean
##          1.8315   -0.8556   -0.4185   -0.4763   5831.4828
## s.e.    0.0574    0.0576    0.1013    0.0950   307.7356
##
## sigma^2 estimated as 16673:  log likelihood = -512.29,  aic = 1036.59
##
## $degrees_of_freedom
## [1] 76
##
## $ttable
##      Estimate      SE  t.value p.value
## ar1      1.8315  0.0574  31.9065  0e+00
## ar2     -0.8556  0.0576 -14.8628  0e+00
## sar1    -0.4185  0.1013  -4.1321  1e-04
## sar2    -0.4763  0.0950  -5.0121  0e+00
## xmean 5831.4828 307.7356  18.9497  0e+00
##
## $AIC
## [1] 12.79736
##
## $AICc
## [1] 12.80723
##
## $BIC
## [1] 12.97472
```

Ainda assim, os gráficos estão longe do White Noise ideal. Contudo, mesmo havendo uma certa variação na média, não existe uma linha de tendência nos resíduos. Já que ele apresenta pouca variação na diferenciação.

Logo, as previsões foram realizadas para um horizonte de 12 trimestres, ou 4 anos.

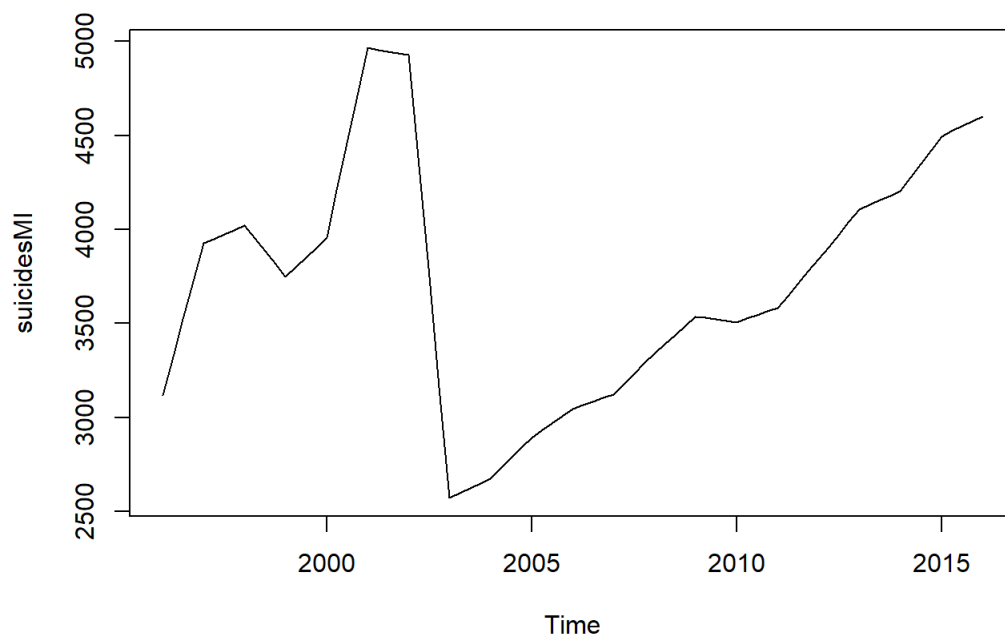


```
## $pred
##      Qtr1      Qtr2      Qtr3      Qtr4
## 2016      7233.649 7154.029 7018.415
## 2017 6836.091 6673.968 6510.192 6349.004
## 2018 6193.965 6064.660 5966.323 5897.824
## 2019 5857.477
##
## $se
##      Qtr1      Qtr2      Qtr3      Qtr4
## 2016      129.1246 269.4431 420.3520
## 2017 572.4306 687.9363 779.3965 852.1287
## 2018 909.3419 942.4989 959.9495 967.6319
## 2019 969.9469
```

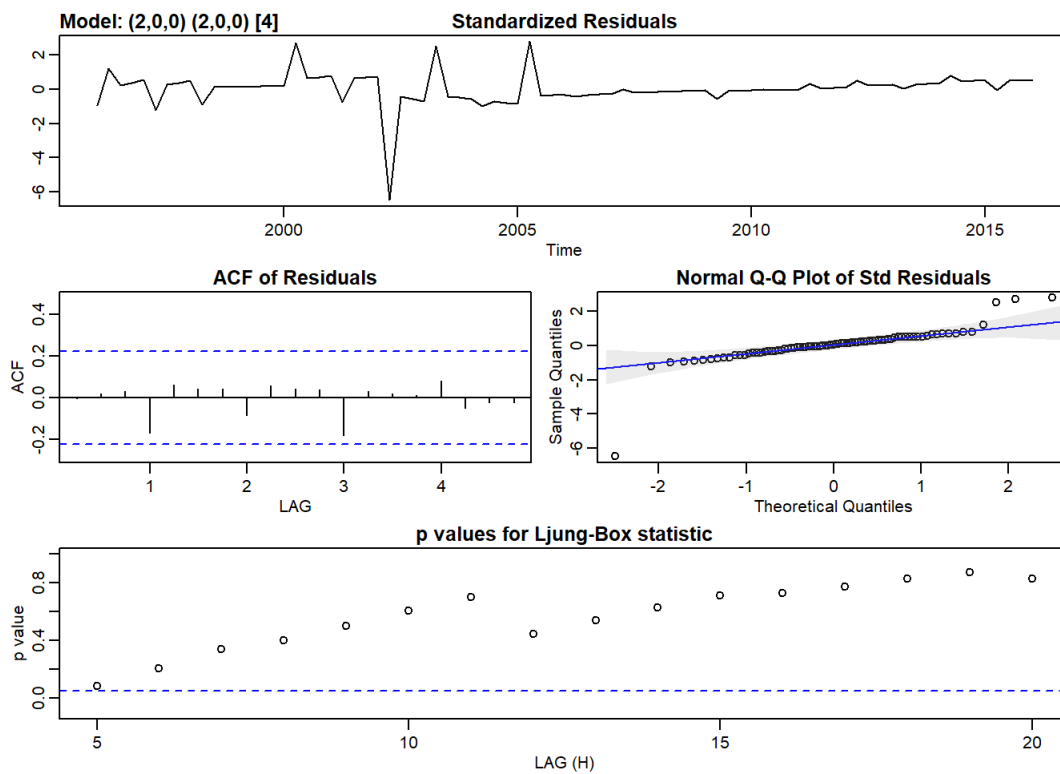
A linha de tendência dos últimos dados está claramente positiva, mas os efeitos passados foram um fator importante para prever a queda dos suicídios. Apesar de ainda estarmos com uma grande margem de erro.

Isso nos leva a questão principal, os grupos de risco estão em perigo? Entenda-se grupos de risco pela faixa da meia-idade. Como visto na análise exploratória

As mesmas análises realizadas para os dados gerais se repetem para a população de meia idade, que segue o mesmo padrão geral. De crescimento nos últimos anos.

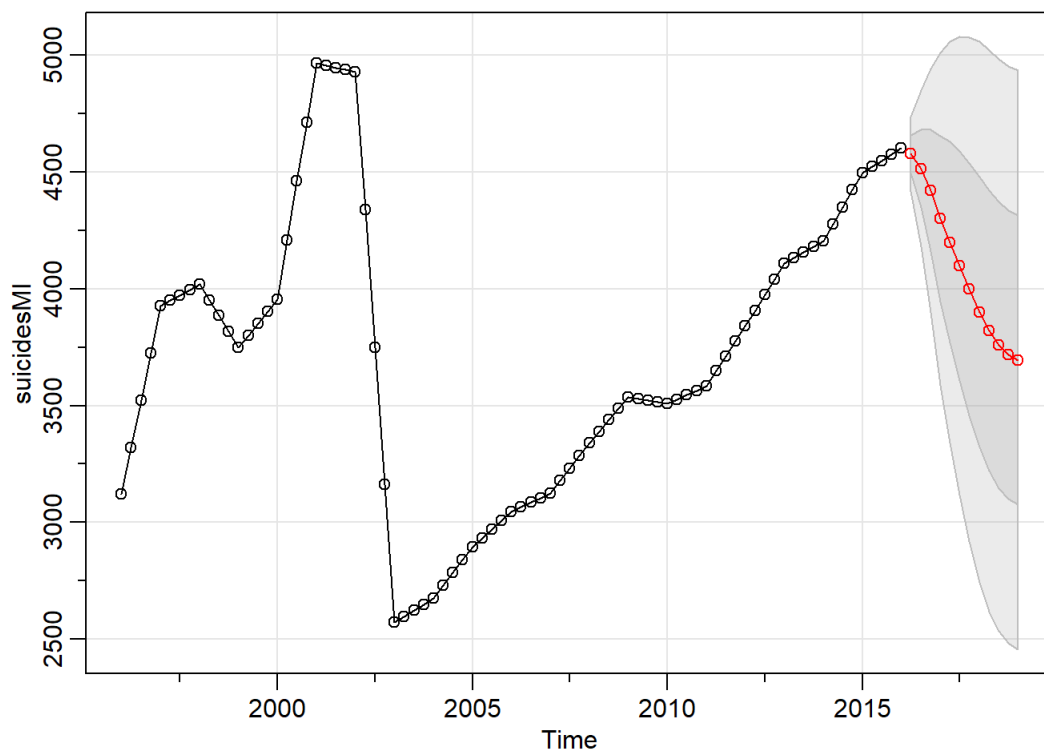


```
## initial value 6.484035
## iter 2 value 6.328429
## iter 3 value 5.421976
## iter 4 value 5.355385
## iter 5 value 5.085991
## iter 6 value 4.972007
## iter 7 value 4.846277
## iter 8 value 4.550493
## iter 9 value 4.471455
## iter 10 value 4.446452
## iter 11 value 4.426286
## iter 12 value 4.416707
## iter 13 value 4.415797
## iter 14 value 4.414473
## iter 15 value 4.414465
## iter 16 value 4.414447
## iter 17 value 4.414446
## iter 18 value 4.414446
## iter 19 value 4.414446
## iter 19 value 4.414446
## iter 19 value 4.414446
## final value 4.414446
## converged
## initial value 4.431589
## iter 2 value 4.431283
## iter 3 value 4.430842
## iter 4 value 4.430714
## iter 5 value 4.430688
## iter 6 value 4.430680
## iter 7 value 4.430670
## iter 8 value 4.430669
## iter 9 value 4.430668
## iter 9 value 4.430668
## iter 9 value 4.430668
## final value 4.430668
## converged
```



```
## $fit
##
## Call:
## stats::arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D,
##     Q), period = S), xreg = xmean, include.mean = FALSE, transform.pars = trans,
##     fixed = fixed, optim.control = list(trace = trc, REPORT = 1, reltol = tol))
##
## Coefficients:
##      ar1      ar2     sar1     sar2    xmean
##      1.8319 -0.8559 -0.3711 -0.4931 3703.149
## s.e.  0.0575  0.0577   0.1007  0.0940 195.549
##
## sigma^2 estimated as 6434:  log likelihood = -473.82,  aic = 959.64
##
## $degrees_of_freedom
## [1] 76
##
## $table
##      Estimate      SE  t.value p.value
## ar1      1.8319  0.0575  31.8761  0e+00
## ar2     -0.8559  0.0577 -14.8227  0e+00
## sar1    -0.3711  0.1007  -3.6840  4e-04
## sar2    -0.4931  0.0940  -5.2464  0e+00
## xmean 3703.1492 195.5490  18.9372  0e+00
##
## $AIC
## [1] 11.84736
##
## $AICc
## [1] 11.85724
##
## $BIC
## [1] 12.02473
```

segundo o mesmo padrão, uma previsão com o mesmo horizonte foi realizada.



```
## $pred
##      Qtr1      Qtr2      Qtr3      Qtr4
## 2016    4577.658  4515.198  4420.512
## 2017  4299.416  4199.313  4097.856  3997.765
## 2018  3901.335  3820.004  3759.233  3718.021
## 2019  3695.042
##
## $se
##      Qtr1      Qtr2      Qtr3      Qtr4
## 2016    80.21446  167.41341  261.23204
## 2017  355.82568  429.86356  489.93166  538.78248
## 2018  578.07612  600.72452  612.59866  617.80523
## 2019  619.36254
```

Conclusão da modelagem

Assim sendo, pode-se concluir que nos últimos anos as taxas de suicídios estão em alta e que, apesar dos modelos atuais preverem uma queda decorrente dos resultados históricos. A margem de confiança ainda está ampla o suficiente para mantermos a atenção nos indicadores de suicídio no Brasil.

Conclusão das Análises

Pode-se interpretar os resultados *estatisticamente* positivos, já que, mesmo com um aumento recente nos suicídios a tendência prevista foi a de uma diminuição. Contudo, pode ser argumentado que as quedas drásticas do número de suicídios em 2003 afetou a qualidade dessa previsão.

Outro ponto relevante foi o impacto que a *política do desarmamento* teve nos suicídios, com uma redução correlacionada ao seu ano de implementação, com o número de suicídios por armas de fogo caindo de forma drásticas. Levando a maior redução nos suicídios masculinos, que utilizam essas ferramentas como uma das principais escolhas no suicídio.

Também foi observado que pessoas entre 25-59 anos cometem significativamente mais suicídios que o resto da população, enquanto recebem menos apoio e políticas direcionadas. Fatores como isolamento social e alienação podem contribuir negativamente para a saúde mental dessas pessoas. Dessa forma, políticas públicas direcionadas a essa faixa etária seriam muito bem vindas.

Analisando as diferenças de gênero, também podemos deduzir que homens tem uma tendência a recorrer a tentativas de suicídios mais “eficazes” como armas de fogo e queda, enquanto mulheres tem uma preferência por envenenamento. Esta diferença também foi considerada como um dos fatores de diferença na taxa entre os gêneros.

Por outro lado, ainda há uma discrepância significativa entre o número de homens e de mulheres que cometem suicídio, com homens tendo quase o dobro de pessoas cometendo o ato. Esse indicador não deve ser ignorado, já que traz insights relevantes na dinâmica social contemporânea.

Dessa forma, as análises conseguiram trazer informações inicialmente contra intuitivas, que desafiaram a hipótese inicial de que jovens e

idosos cometem mais suicídios, e que os gêneros não apresentam diferenças relevantes entre si.