

# Introdução

## Contexto

A análise exploratória de dados a seguir se insere no contexto da pandemia da Covid-19. A Organização Mundial da Saúde (OMS) declarou, em 30 de janeiro de 2020, que o surto da doença causada pelo novo coronavírus constitui uma Emergência de Saúde Pública de Importância Internacional – o mais alto nível de alerta da Organização. Até 31 de julho de 2020 já haviam sido confirmados 17.106.007 casos de COVID-19 e 668.910 mortes. Diferentes ações vem sendo realizadas por cada ente da Federação no combate à doença devido ao caráter inédito da situação, uma vez que não se sabe ao certo os meios mais eficazes de combate. Como destaque, pode-se mencionar as medidas de isolamento social por meio da restrição do comércio, mantendo-se abertos apenas os serviços essenciais. Por outro lado, existem municípios descrentes quanto a essa medida, principalmente devido aos impactos econômicos gerados. Dessa forma, há localidades em que não houveram restrições na atividade econômica. Nesse sentido, uma análise exploratória dos dados que existem referentes à evolução da doença no Brasil se faz necessária, de forma a gerar insumos informacionais capazes de melhor orientar as ações governamentais. Ainda, permite que se compreenda melhor as tendências que podem ser observadas por meio da evolução dos casos nas diferentes regiões, estados e municípios.

## Objetivo

A seguinte pesquisa busca identificar a evolução do número de casos e mortes por covid-19 nos diferentes municípios, estados e regiões do Brasil. Ainda, busca identificar os municípios, estados e regiões com o maior número de casos e de óbitos, analisando se há coincidência entre eles. Será feita análise temporal, por meio dos números obtidos nos meses de maio e junho. Para além dos dados, busca identificar quais medidas foram tomadas nos municípios que tiveram maior destaque, tanto positivo quanto negativo, no combate à proliferação da doença. Para tanto, será utilizada a base de dados referente aos casos da doença no país disponibilizado pelo portal Brasil.io (<https://brasil.io/dataset/covid19/caso/>).

## Perguntas

1. Quais os municípios com o maior número de casos confirmados de Coronavírus?
2. Quais os municípios com o maior número de mortos por Coronavírus?
3. Os municípios com os maiores números de casos confirmados são os mesmos com os maiores números de mortos?
4. Fazer as mesmas análises para Estados e Regiões do Brasil.
5. Quais os 5 estados e 5 municípios que mais aumentaram o número de casos no mês de junho com relação ao maio?
6. Quais os 5 estados com o menor crescimento de casos no mês de junho com relação ao maio?
7. Quais os 5 estados e 5 municípios que mais aumentaram o número de óbitos no mês de junho com relação a maio?
8. Como foi a evolução dos casos nos estados e nos municípios identificados nas questões acima?

## Descrição dos dados

A base original consiste em um dataframe contendo 386665 observações com 17 variáveis. As principais variáveis que nos interessam nessa base são o município, a data, os novos casos confirmados, as novas mortes, o estado, e o código da cidade. Como a base em suas linhas contém informações tanto de estados quanto de municípios, esse é um ponto que precisaremos sanar na arrumação dos dados, que segue abaixo

## Base original

```
df <- read.csv("./Dados/caso_full.csv")
kable(head(df))
```

city	city_ibge_code	date	epidemiological_week	estimated_population_2019	is_last	is_repeated
São Paulo	3550308	2020-02-25	9	12252023	False	False
	35	2020-02-25	9	45919049	False	False
São Paulo	3550308	2020-02-26	9	12252023	False	False
	35	2020-02-26	9	45919049	False	False
São Paulo	3550308	2020-02-27	9	12252023	False	False
	35	2020-02-27	9	45919049	False	False

## Organizando os dados

Selecionando apenas as cidades, depois selecionando apenas as colunas de interesse e mudando o tipo da coluna de códigos do IBGE para alterarmos os nomes das cidades posteriormente. Por fim adicionamos uma coluna com os meses, separados das datas, pois é uma informação que usaremos para responder nossas perguntas.

```
df <- df %>% filter(place_type == "city") %>%
select(c("date",
         "city",
         "new_confirmed",
         "new_deaths",
         "state",
         "estimated_population_2019",
         "epidemiological_week",
         "city_ibge_code")) %>%
mutate(city_ibge_code = as.character(city_ibge_code)) %>%
mutate(Mes = substring(date,6,7))
```

Trazendo tabela com os códigos do IBGE e nomes certos das cidades

```
codes <- read_excel("./Dados/Codigo_IBGE.xls")
```

Trazendo o nome correto para o Dataframe

```
dfarrumado <- left_join(df,codes,by = c("city_ibge_code"="Código"))
```

Excluindo a coluna com os nomes mal formatados

```
dfarrumado$city <- NULL
```

Arrumando a ordem e o nome das colunas

```
df <- dfarrumado[,c(1,8,7,9,4,5,2,3)]
colnames(df) <- c("data", "mes", "codigo", "municipio", "estado", "popest2019", "novoscasos", "novasmortes")
df$popest2019 <- NULL
```

## Tabela gerada

Apenas as 6 primeiras linhas

data	mes	codigo	municipio	estado	novoscasos	novasmortes
2020-02-25	02	3550308	São Paulo	SP	1	0
2020-02-26	02	3550308	São Paulo	SP	0	0
2020-02-27	02	3550308	São Paulo	SP	0	0
2020-02-28	02	3550308	São Paulo	SP	1	0
2020-02-29	02	3550308	São Paulo	SP	0	0
2020-03-01	03	3550308	São Paulo	SP	0	0

Observando algumas informações do data.frame

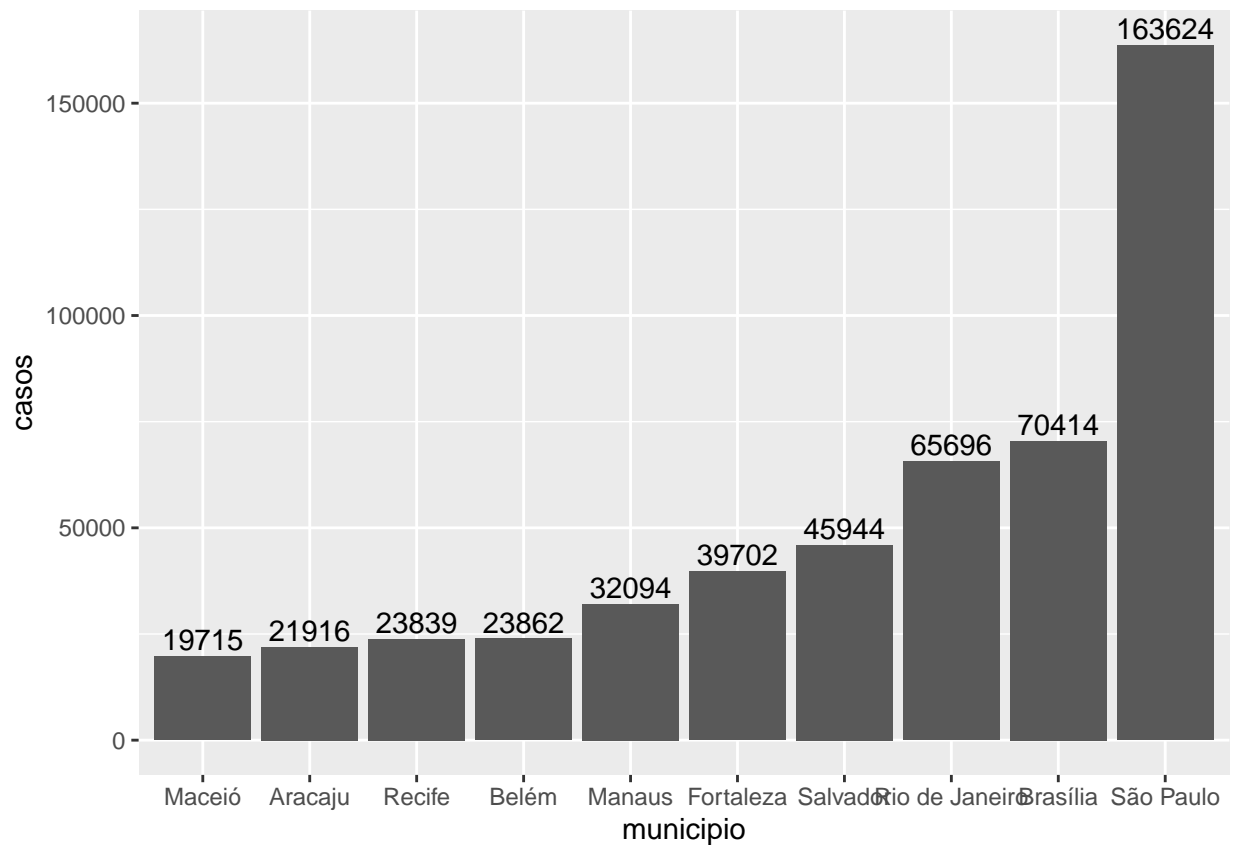
```
##      data      mes      codigo      municipio
## Length:383205 Length:383205 Length:383205 Length:383205
## Class :character Class :character Class :character Class :character
## Mode  :character Mode  :character Mode  :character Mode  :character
##
##
##      estado      novoscasos      novasmortes
## Length:383205 Min.    :-1578.000 Min.    :-82.0000
## Class :character 1st Qu.:   0.000 1st Qu.:  0.0000
## Mode  :character Median :   0.000 Median :  0.0000
##                Mean   :   5.353 Mean   :  0.2033
##                3rd Qu.:   2.000 3rd Qu.:  0.0000
##                Max.   : 6877.000 Max.   : 227.0000
```

podemos observar como informações mais interessantes o fato do máximo de novos casos em um município em um dia ter sido 6877 novos casos e o máximo de mortes em um único município em um dia ter sido de 227 óbitos

## Análise Exploratória

### 1. Quais os municípios com o maior número de casos confirmados de Coronavírus?

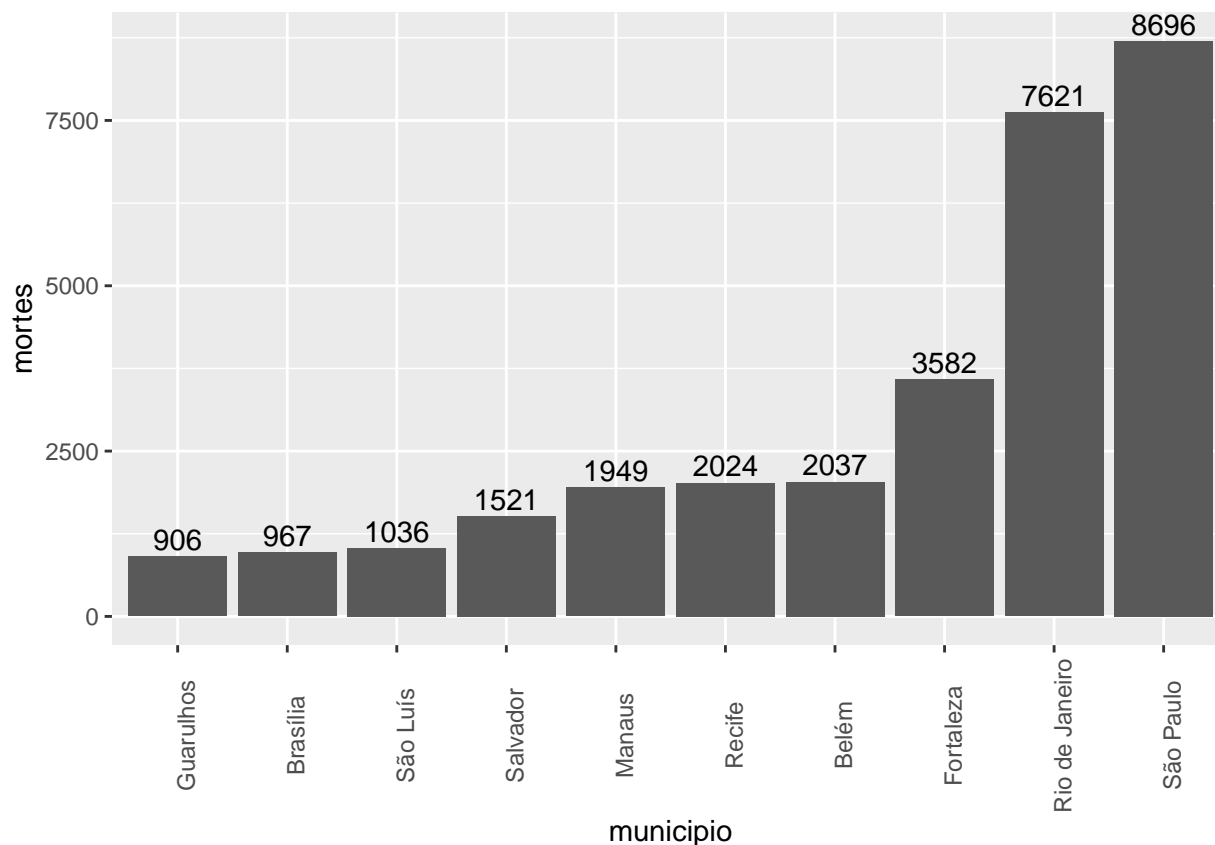
Aqui optamos por fazer um simples gráfico de barras com os 10 municípios com mais casos até 18/07/2020



Percebemos que os 10 municípios que tiveram mais casos até o dia 18 de julho de 2020 foram São Paulo, Brasília e Rio de Janeiro. destaca-se a diferença gritante entre o primeiro e segundo lugar, em que São Paulo tem cerca de 132% a mais de casos que Brasília.

## 2. Quais os municípios com o maior número de mortos por Coronavírus?

Aqui seguimos a mesma linha para responder a pergunta anterior.

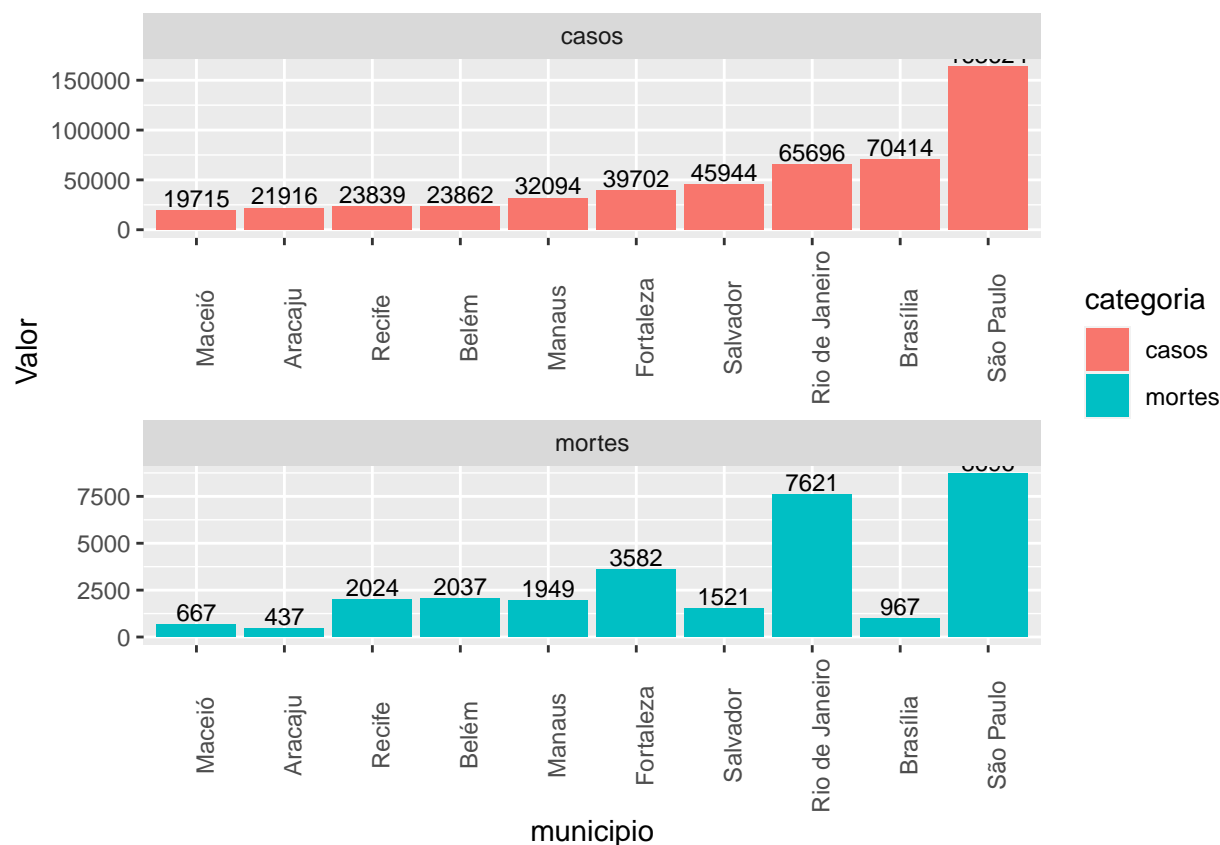


Com a análise deste gráfico é possível observar que o município de São Paulo lidera tanto no quesito número de casos confirmados, quanto no quesito número de mortes confirmadas por Coronavírus. As medidas adotadas pelo governo paulista para o enfrentamento do avanço da doença visaram a redução da taxa de contágio e incluíram desde a criação de um plano de ação com a destinação de R\$200 mil para aquisição de kits de diagnósticos, a criação do Centro de Contingência do Coronavírus para monitorar e coordenar as ações contra a propagação da doença e a criação do Comitê Administrativo Extraordinário, com o poder de tomar medidas de caráter emergencial durante a pandemia. Na esfera municipal, a prefeitura de São Paulo instituiu o regime de isolamento social, com o fechamento do comércio, funcionando apenas serviços essenciais, como supermercados, farmácias e postos de gasolina e também instituiu um plano de ação para o enfrentamento da doença. A alta taxa de casos e mortes na cidade se justifica por sua grande população de mais de 21,5 milhões de habitantes incluindo a região metropolitana.

Entretanto, cabe aqui ressaltar alguns casos em que há grande discrepância entre o número de casos e o número de óbitos por essa doença. Sendo assim, destaca-se o caso de Brasília, que apesar de ser o segundo município com maior número de casos - superando até mesmo o município do Rio de Janeiro que tem recebido grande atenção pelos noticiários -, encontra-se apenas na 9ª colocação no que se refere ao número de óbitos confirmados, cabendo uma possível reflexão acerca da maior eficiência do atendimento médico-hospitalar de Brasília, ou um possível atraso na confirmação dos óbitos ocorridos pela doença. Dessa forma, tendo em vista a existência dessa discrepância entre número de casos e número de mortos causados pela Covid-19, no gráfico seguinte buscar-se-á analisar essa questão mais profundamente.

### 3. Os municípios com os maiores números de casos confirmados são os mesmos com os maiores números de mortos?

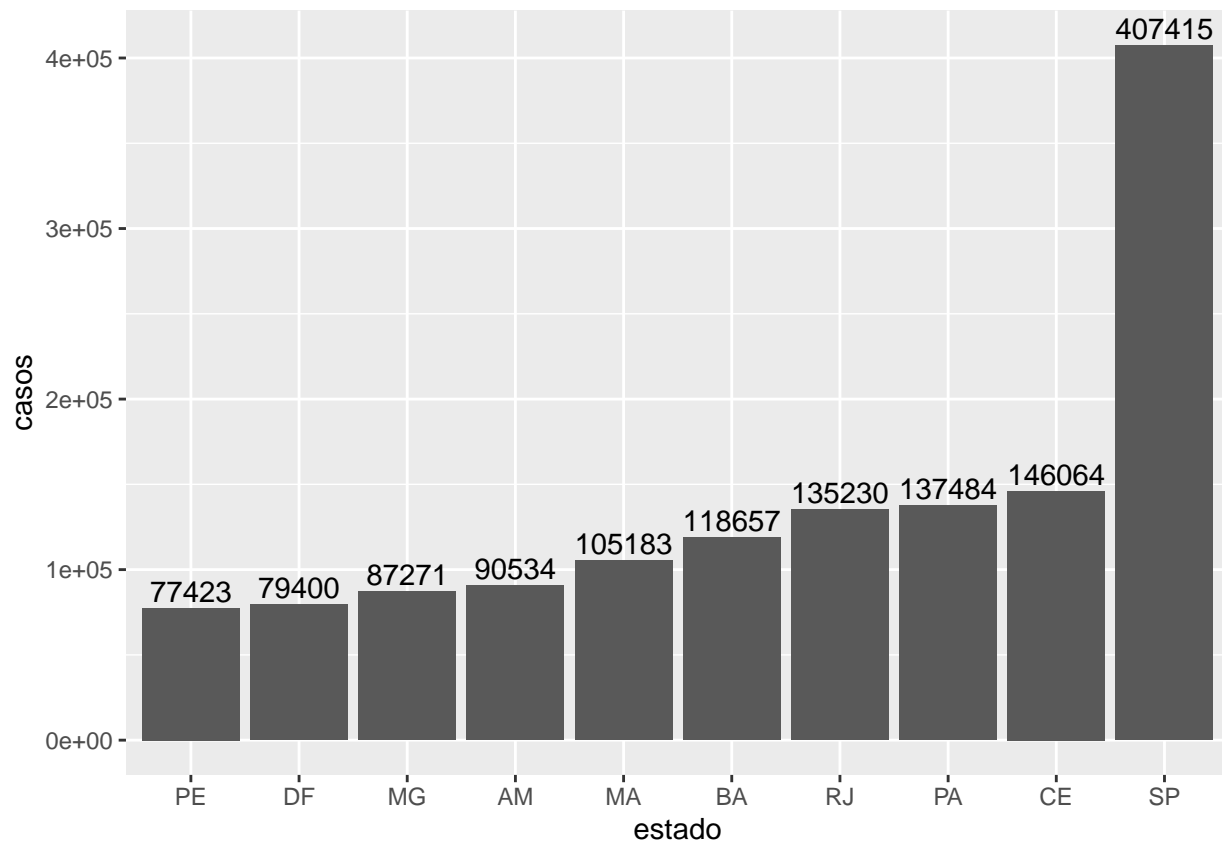
Aqui podíamos apenas comparar as duas últimas visualizações, mas optamos por gerar uma visualização com ambos os gráficos, de modo a facilitar a comparação.



Com este gráfico, a análise feita anteriormente fica ainda mais clara. É possível notar que alguns municípios têm se destacado positivamente no quesito número de óbitos ocorridos, como é o caso de Brasília, de Salvador e de Manaus, que apesar de um relativo elevado número de casos, apresentou um relativo menor número de óbitos. Brasília, por exemplo, sancionou a lei 6.552/2020, trazendo uma série de medidas para redução dos riscos da doença e a garantia de acesso aos serviços de prevenção, tratamento e cura da Covid 19. Essa lei previu o fornecimento de equipamentos respiratórios e de proteção individual adequados e necessários aos agentes de saúde, limitação do tráfego local de pessoas e veículos, a transparência na coleta, tratamento e divulgação dos casos de contaminados, doentes e mortos. Ademais, equipes da Secretaria de Proteção e Ordem do Distrito Federal, em parceria com a Polícia Militar, realizam diversas ações de fiscalização e orientação a consumidores e trabalhadores em feiras, vistorias do uso de máscaras e cumprimento de medidas sanitárias de combate ao vírus. Por outro lado, há alguns municípios que se destacam negativamente neste mesmo quesito, tendo em vista que apesar de um relativo menor número de casos, apresentou um relativo maior número de óbitos, como é o caso do Rio de Janeiro e de Fortaleza. O Rio de Janeiro tomou medidas como a inauguração de 13 clínicas da família, voltadas para ações de prevenção, promoção da saúde e diagnóstico, a compra e entrega de equipamentos para modernização tecnológica de hospitais, manteve as escolas fechadas e as medidas de isolamento social, a aplicação de multas para quem for flagrado sem máscara e interrompeu festas e aglomerações em diversos estabelecimentos.

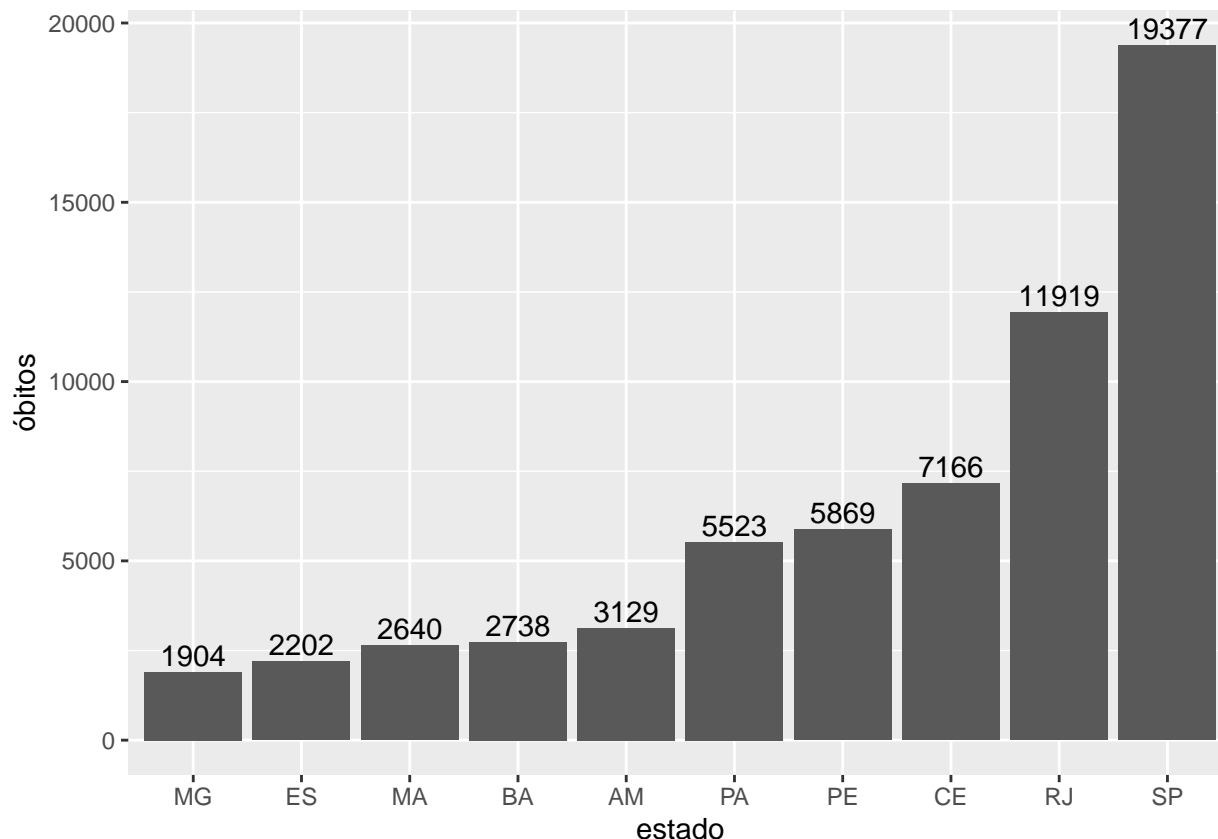
#### 4. Fazer as mesmas análises para Estados do Brasil.

Aqui seguimos a mesma lógica das perguntas anteriores, porém substituindo municípios por estados. Abaixo gráfico de casos



Neste gráfico, é possível observar a grande discrepância entre o primeiro estado com maior número de casos confirmados (São Paulo) e os demais 9 estados com maiores números de casos confirmados. Nesse sentido, cabe ainda destacar que a partir do segundo estado com maior número de casos confirmados, não há discrepâncias tão significativas entre cada um desses 9 estados. Verificou-se no gráfico que Pernambuco foi o estado com o menor número de casos entre os 10 com mais casos. O governo de Pernambuco instituiu uma portaria que dispõe sobre o funcionamento de estabelecimentos com serviços de alimentação e recomendações de medidas preventivas, suspendeu as aulas na rede pública e privada, suspendeu consultas, procedimentos diagnósticos terapêuticos ambulatoriais e também cirurgias eletivas. O governo instituiu também, regras de distanciamento e isolamento social, além de rodízio de veículos de acordo com o número da placa.

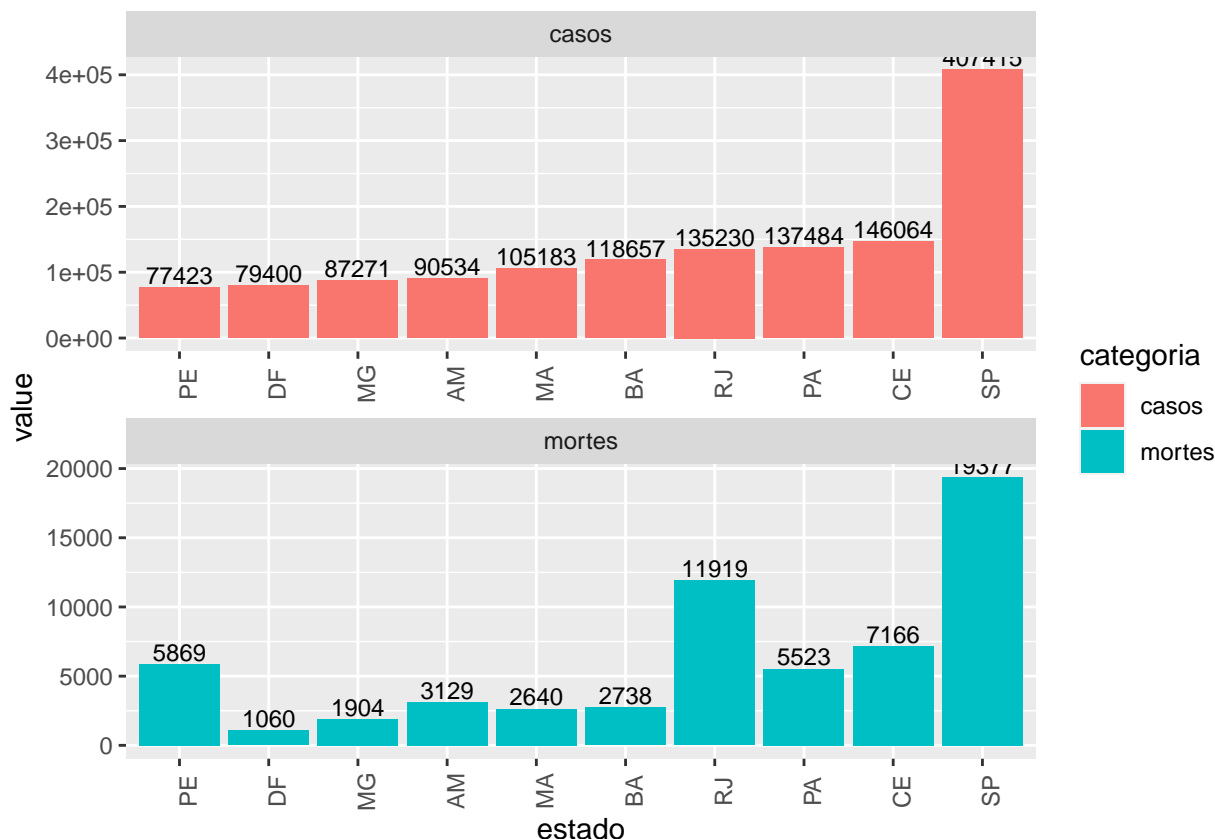
Abaixo gráfico de óbitos:



Neste gráfico, por sua vez, é interessante destacarmos que o primeiro estado com o maior número de óbitos confirmados corresponde ao estado com maior número de casos confirmados, como era de se esperar. Entretanto, nota-se, também, algumas discrepâncias, como por exemplo o fato de o estado do Rio de Janeiro ser o segundo estado com maior número de óbitos confirmados, apesar de ele se encontrar apenas em quarto lugar no que se refere ao número de casos confirmados. Percebemos também no gráfico acima que Minas Gerais foi o estado com menor número de óbitos entre os 10 estados de maior número de mortes. O governo mineiro tem dado atenção especial ao momento de pandemia, através de medidas de isolamento social, acompanhamento diário dos casos, ampliação de leitos hospitalares, suspensão de cirurgias eletivas, compra de equipamentos de EPI para proteção dos profissionais da saúde que estão atuando na linha de frente do combate ao vírus no estado. Sendo assim, o gráfico a seguir dará maior suporte a essas e outras análises comparativas entre os estados com maior número de casos confirmados e com maior número de óbitos confirmados.

E por fim a comparação:





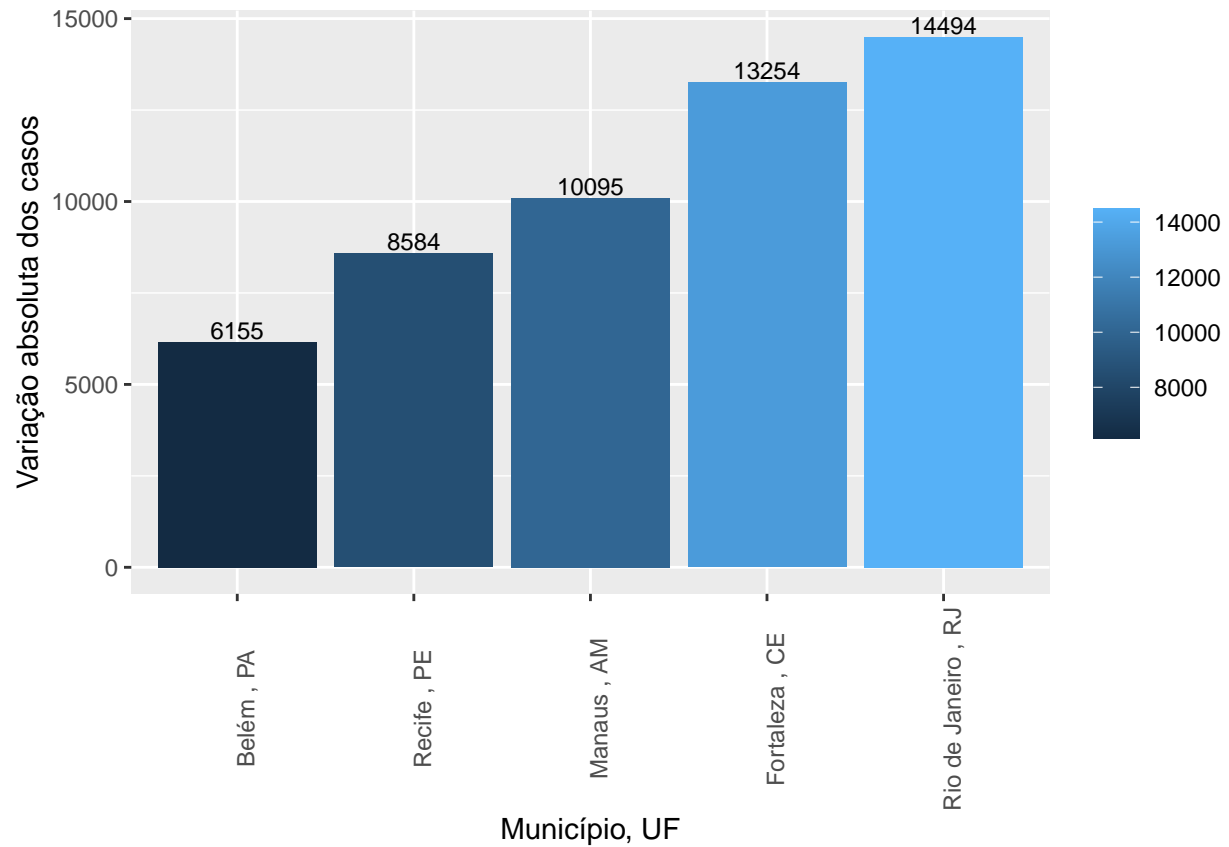
Neste gráfico, assim como no gráfico 3, a análise comparativa entre os estados com maior número de casos confirmados e os estados com maior número de óbitos confirmados fica ainda mais evidente. Sendo assim, cabe aqui destacar negativamente o caso do Rio de Janeiro, de Pernambuco e do Amazonas, tendo em vista que apesar de apresentarem um relativo menor número de casos confirmados, apresentam um maior número de óbitos. Por outro lado, destaca-se positivamente o estado do Maranhão, que apesar de apresentar um relativo maior número de casos confirmados, apresentou um relativo menor número de óbitos. O governo do Maranhão vem atuando no combate ao coronavírus, através de vários decretos que estabelecem medidas preventivas e restritivas, como a suspensão do ensino presencial, regras de funcionamento das atividades econômicas, medidas de isolamento social e prevenção do contágio e propagação do coronavírus.

**5. Quais os 5 estados e 5 municípios que mais aumentaram o número de casos no mês de junho com relação ao maio?**

**6. Quais os 5 estados com o menor crescimento de casos no mês de junho com relação ao maio?**

Aqui vimos que uma visualização para casos e morte no contexto do estados já nos permitira identificar aqueles com maior e menor crescimento. Optamos também por duas estratégias para averiguação da variação: a variação absoluta e a variação relativa. Abaixo segue o gráfico do municípios que tiveram o maior crescimento absoluto entre maio e junho:

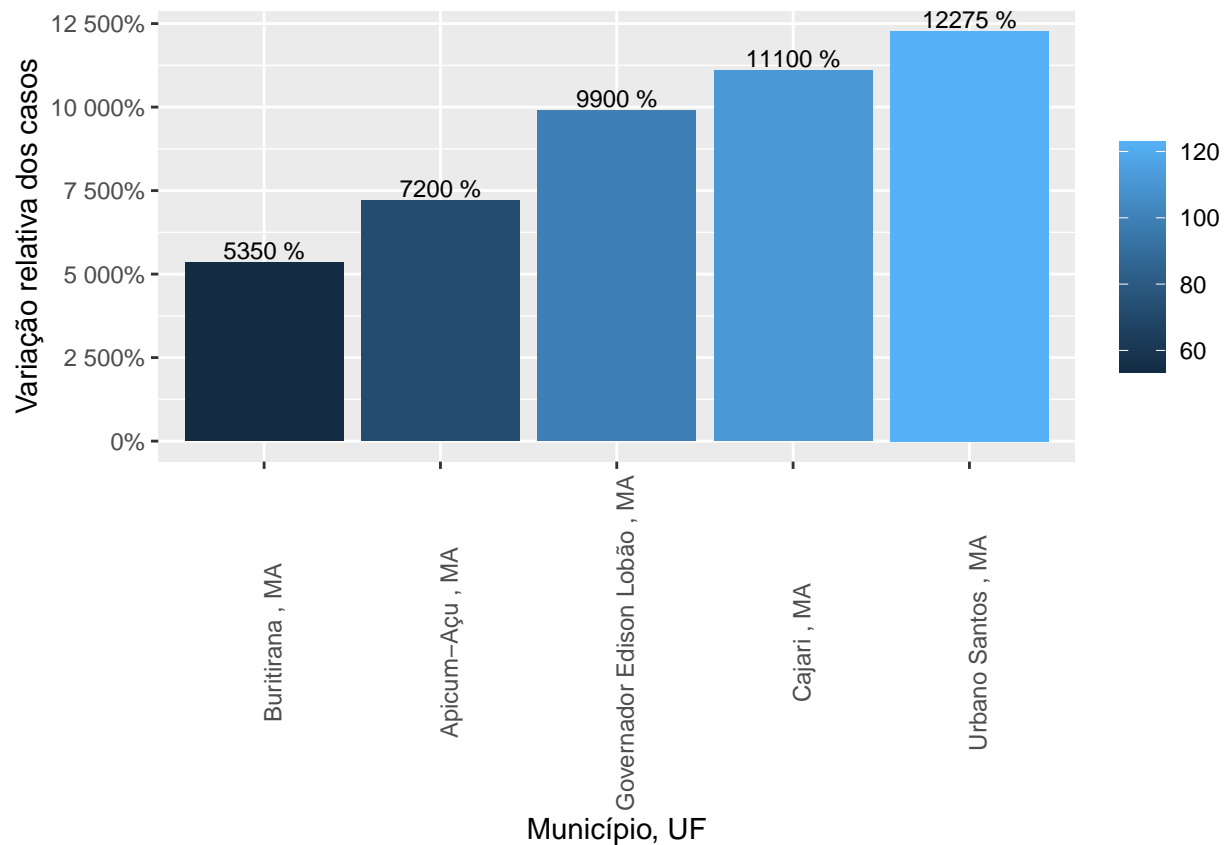
```
## Adding missing grouping variables: 'estado', 'municipio'
```



Neste gráfico, é interessante notarmos que todos os 5 municípios que apresentaram maior crescimento absoluto no número de casos entre maio e junho são capitais de seus respectivos estados

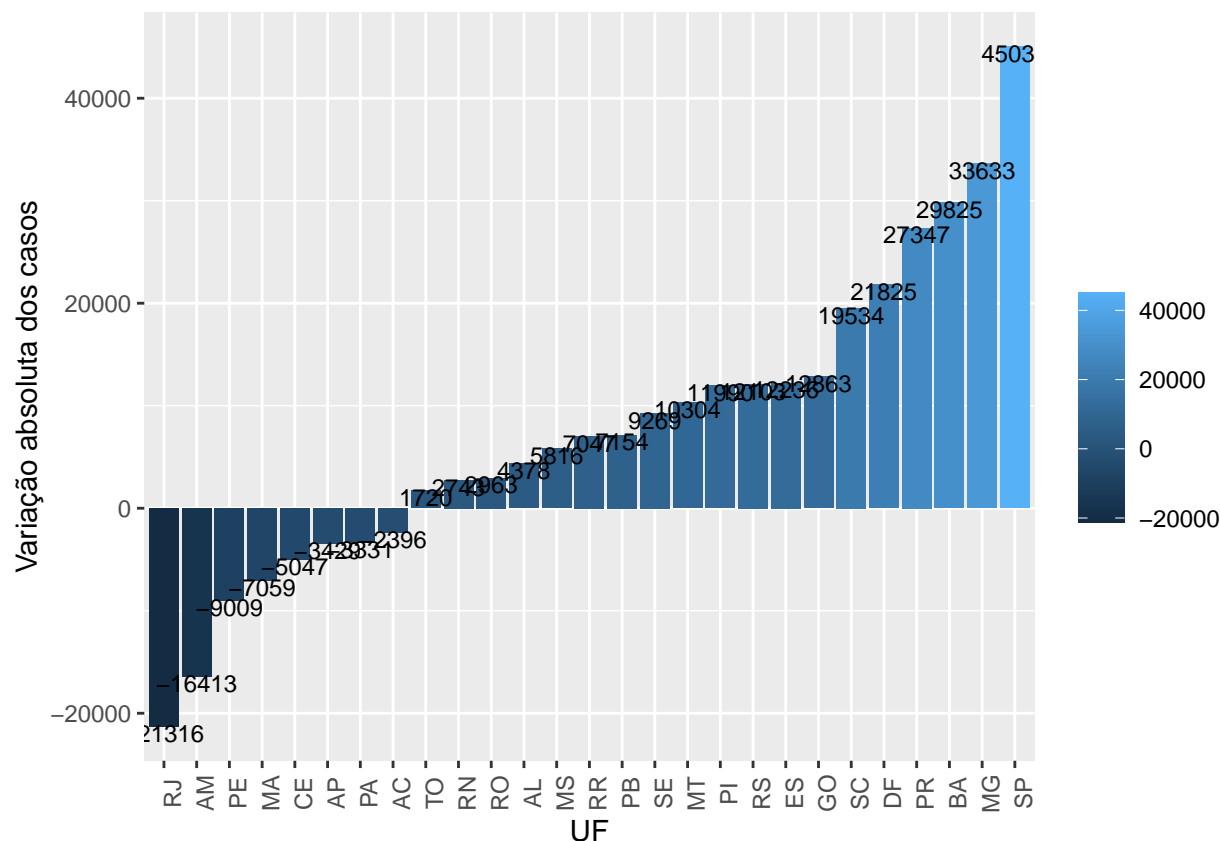
Em seguida, os municípios com maior crescimento relativo. Para gerar esse gráfico optamos por excluir aqueles municípios que não tinham nenhum caso em maio.

```
## Adding missing grouping variables: 'estado', 'municipio'
```

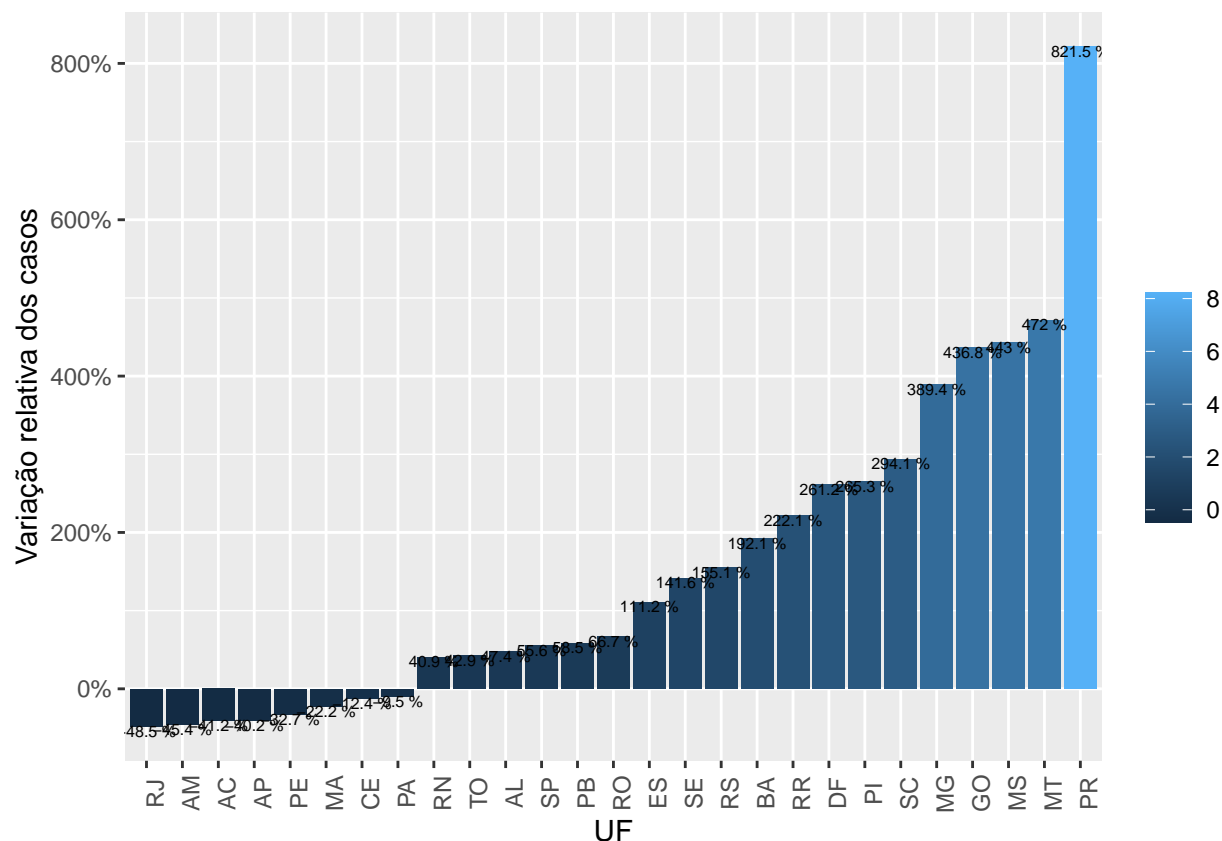


Neste gráfico é possível observar que os municípios que apresentaram maior crescimento relativo no número de casos entre maio e junho não são os mesmos que apresentaram maior crescimento absoluto, tendo em vista, basicamente, uma questão estatística, tendo em vista que municípios que apresentavam menos de 5 casos em maio e saltaram para centenas de casos em junho, apresentam um crescimento relativo muito maior que os municípios que já apresentavam centenas ou até mesmo milhares de casos em maio.

Abaixo temos o gráfico com a variação absoluta dos estados:



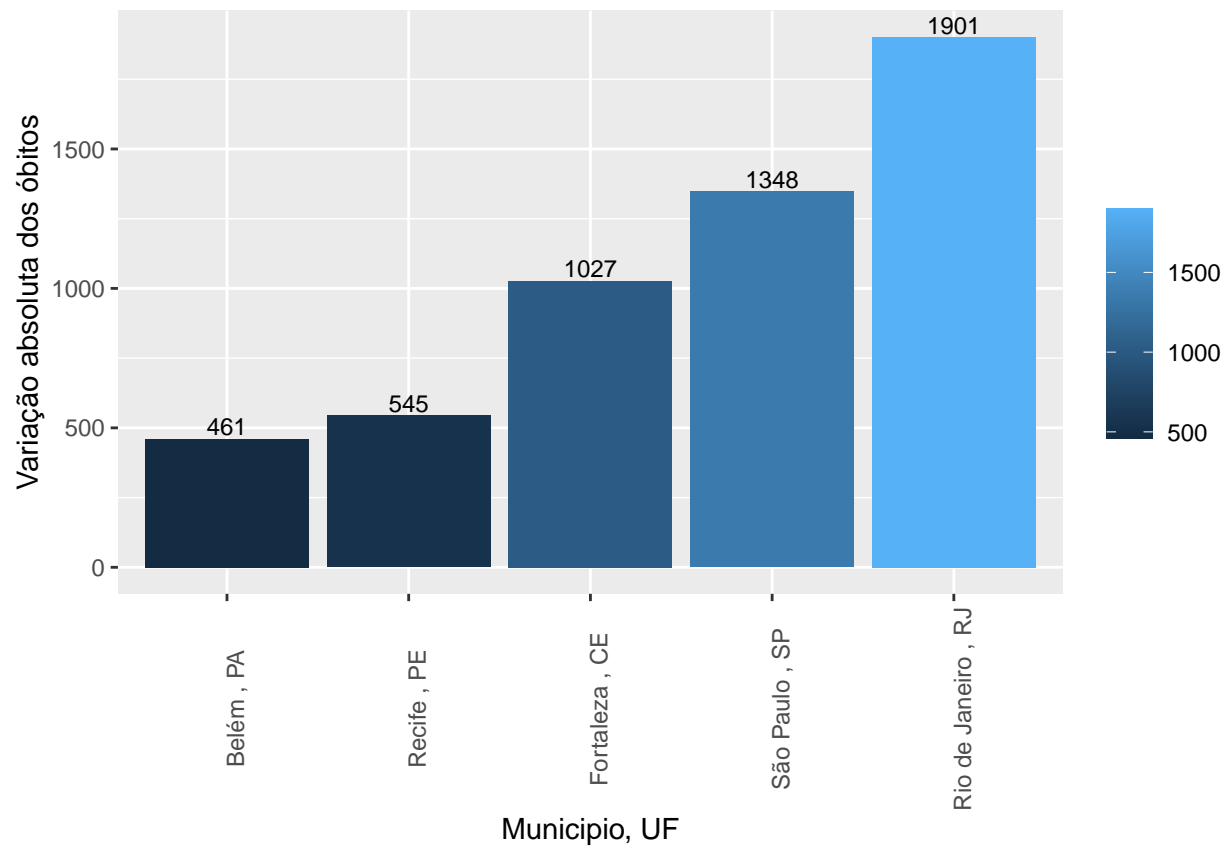
Neste gráfico, cabe destacar que a existência de uma variação absoluta negativa do número de casos significa que entre o mês de maio e junho houve alguns estados em que se observou a confirmação de menos casos no mês de junho do que se observou em maio. Ou seja, em alguns estados, como no Rio de Janeiro e no Amazonas, por exemplo, houve uma redução no crescimento da doença. Por outro lado, em alguns estados, como São Paulo e Minas Gerais, por exemplo, é possível observar uma variação absoluta dos casos muito positiva, fazendo com que houvesse um crescimento no número de contaminados pela doença. O governo do estado do Amazonas implantou mais de 140 ações de enfrentamento a pandemia da Covid 19, que vão desde ações para a desaceleração da transmissão da doença até apoio à atividade econômica e fomento à pesquisa e por fim, a variação relativa dos estados



Neste gráfico é interessante observarmos que os Estados com maior variação relativa dos casos não são os mesmo com maior variação absoluta dos casos, tendo em vista a mesma questão estatística supracitada. Dessa forma, é interessante ressaltarmos também a grande variação relativa dos casos no Paraná, demonstrando grande crescimento do número de casos confirmados no estado. O Paraná adotou medidas de enfrentamento ao coronavírus através de decreto do governador, no sentido de reduzir as possibilidades de contato entre as pessoas para conter a propagação do vírus, além de ações específicas no âmbito da saúde, ampliação da proteção de servidores incluídos no grupo de risco e medidas de fortalecimento no controle das fronteiras. Uma das medidas do governo paranaense foi a suspensão de férias e licenças de servidores das secretarias de saúde, segurança pública e da Coordenadoria Estadual de Proteção e Defesa Civil, objetivando a garantia de profissionais dessas áreas durante o surto.

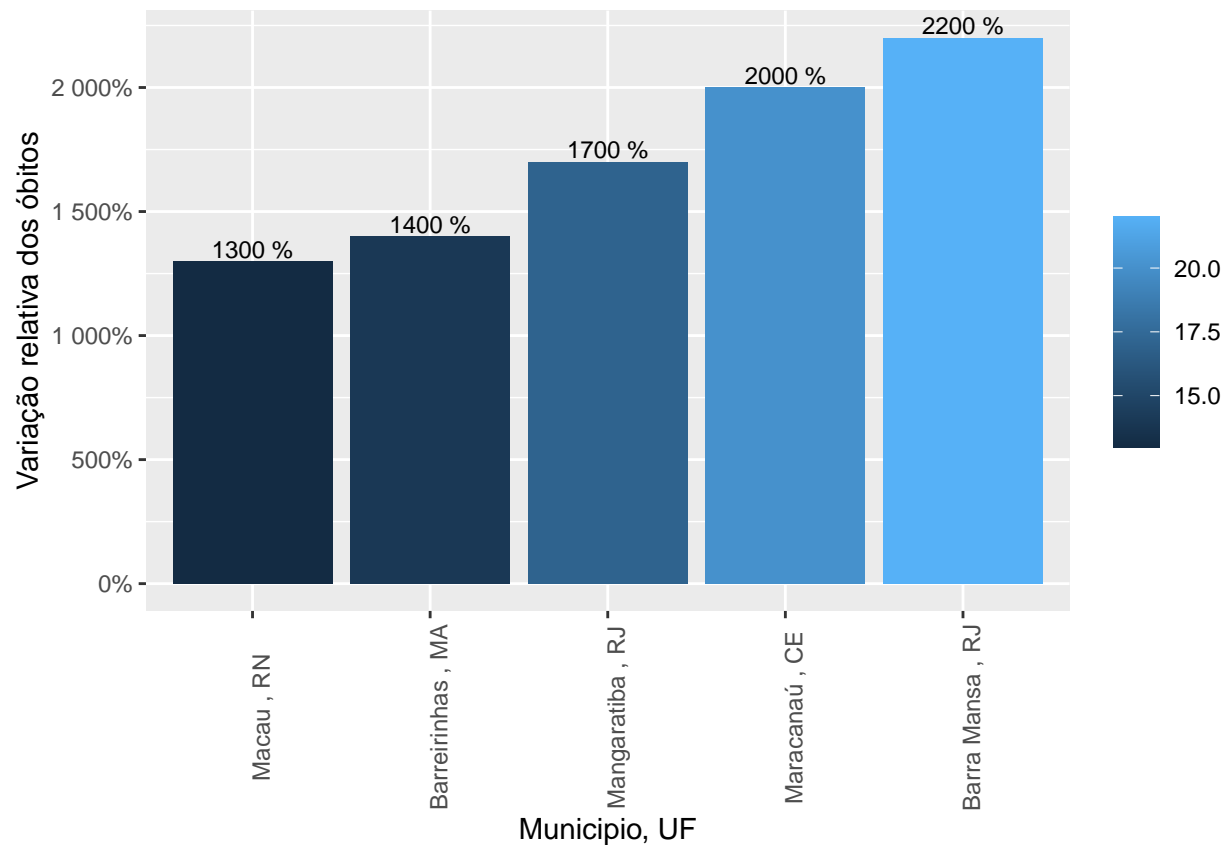
## 7. Quais os 5 estados e 5 municípios que mais aumentaram o número de óbitos no mês de junho com relação a maio?

Aqui seguimos a mesma lógica da questão anterior, só que averiguando o número de óbitos. Abaixo temos o gráfico com a variação absoluta de mortes por município



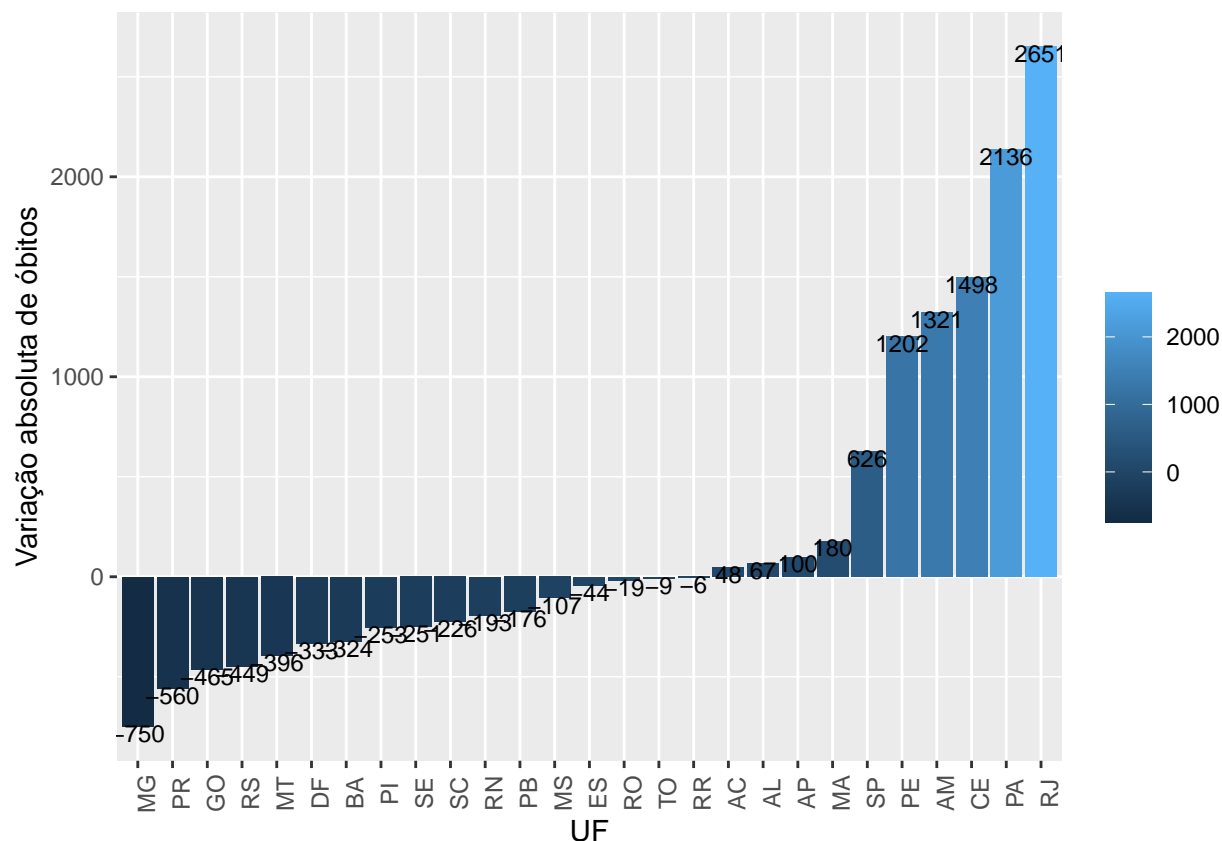
– Neste gráfico, novamente, é interessante notarmos que os 5 municípios que apresentaram maior variação absoluta no número de óbitos no mês de junho com relação a maio são todos capitais de seus respectivos estados.

Em seguida observamos a variação relativa:



Neste gráfico, é interessante analisarmos, assim como foi feito no gráfico correspondente aos municípios com maior variação relativa dos casos entre maio e junho, que os municípios com maior variação absoluta dos óbitos nesse mesmo período, não são os mesmos que obtiveram maior variação relativa dos óbitos.

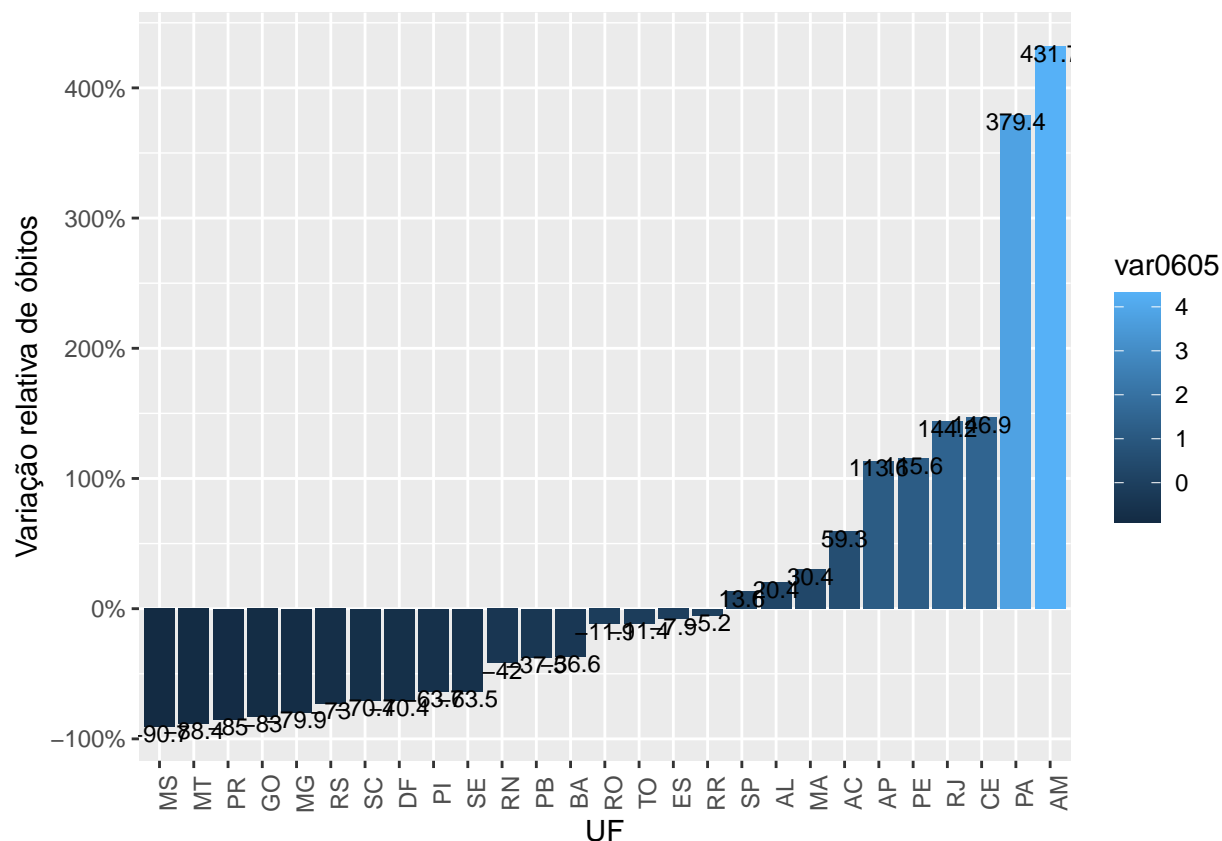
Obtivemos então o gráfico da variação absoluta dos estados:



Nesse gráfico é interessante notarmos alguns pontos. Em primeiro lugar o fato de alguns estados terem apresentado variação absoluta de óbitos entre maio e junho negativa, demonstrando que houve certa diminuição no número de mortos por Coronavírus dentro do período analisado, destacando-se o estado de Minas Gerais e do Paraná. Por outro lado, destaca-se também que alguns estados apresentaram variação absoluta de óbitos entre maio e junho positiva, demonstrando que ocorreu um aumento no número de mortos por Coronavírus em alguns estados, destacando-se o Para e o Rio de Janeiro, que apresentaram no mesmo período uma variação absoluta negativa no número de casos. O governo do Pará montou, ainda em janeiro, um plano de contingência com o objetivo inicial de preparar o estado para o enfrentamento da Covid 19. Foi realizado um fortalecimento da rede hospitalar com a abertura de novos hospitais, aumentando 720 novos leitos na rede pública, instalação de hospitais de campanha, adquiriu 400 novos respiradores dobrando a capacidade das unidades de terapia intensiva, reduziu os gastos públicos congelando despesas que não fossem referentes a saúde e determinou que não fossem feitos novos contratos, com o intuito de trazer tranquilidade financeira durante o enfrentamento do vírus.

e por fim o de variação relativa:

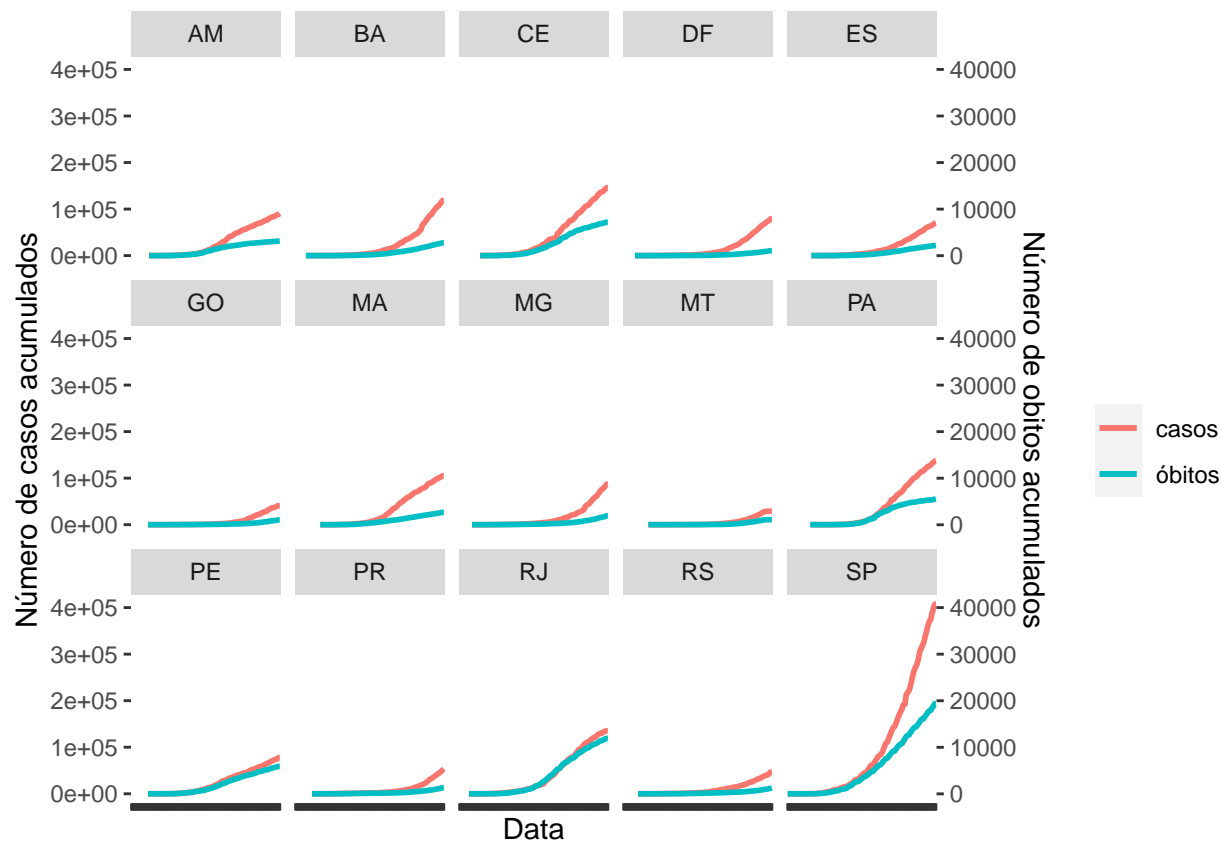




Neste gráfico é interessante notarmos que os estados com maior variação relativa de óbitos não são os mesmo que apresentaram maior variação absoluta de óbitos, tendo em vista, mais uma vez, a questão estatística citada anteriormente.

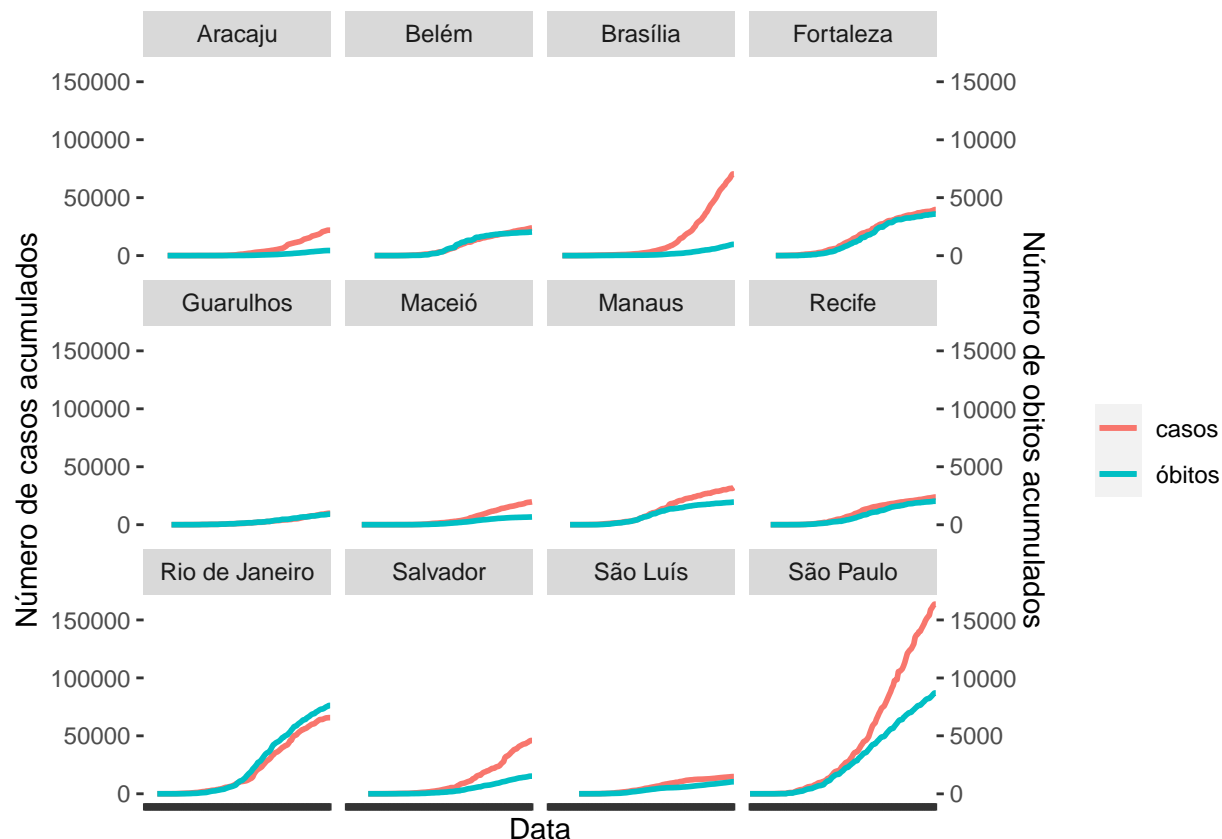
## 8. Como foi a evolução dos casos nos estados e nos municípios identificados nas questões acima?

Para responder essa pergunta procuramos todos os municípios e estados que apareceram nas explorações anteriores, de modo que fizéssemos apenas duas visualizações: 1 com o estados que apareceram e outra com os municípios. Nelas também colocamos a serie temporal dos óbitos, porém adicionando um eixo y secundário para que seja possível observarmos a variação. Essas visualizações podem ser observadas a seguir:



Este gráfico é interessante para que possamos comparar os níveis de desenvolvimento da doença, tanto no que se refere ao número de casos confirmados, quanto ao número de óbitos confirmados dos 15 estados mais impactados pelo novo Coronavírus. Sendo assim, destacam-se os casos do Ceará, Minas Gerais, Pará, Rio de Janeiro, e São Paulo, que tem apresentado uma evolução notável da doença. O governo cearense, por sua vez, também tomou medidas para o enfrentamento à pandemia através do isolamento social, fechamento do comércio, igrejas, museus e barracas de praia, além de fiscalização sanitária nas divisas do estado através de barreiras terrestres nas rodovias, com auxílio da Polícia Rodoviária Estadual.

Abaixo os gráficos dos municípios:



Por fim, este gráfico é importante para que possamos ter um panorama geral dos impactos da doença nas 15 capitais mais impactadas pelo novo Coronavírus. Cabe destacar, portanto, as capitais Brasília, Rio de Janeiro e São Paulo, que tem suas curvas de Casos e de Óbitos mais ascendentes.

## Limitações e dificuldades operacionais

As principais limitações de nossa análise foi o fato de não considerarmos a população dos municípios, não calculando coeficientes de mortalidade para os municípios e estados que analisamos. Pela definição das perguntas ter sido prévia, e no momento que as definimos não incluímos análises incluindo as populações, acabamos não inserindo essa variável na análise. Como os dados estavam um pouco desorganizados, a fase de arrumação foi bem desafiadora, mas, com os dados arrumados, a geração das visualizações foi bem tranquila. A utilização dos pipes trouxe imensa produtividade ao processo.