

Curso de Vigilância em Saúde apoiado por plataforma BI

Aula 04 – Calculando taxas de incidência temporais

1. Apresentação

Olá!

Seja bem-vindo à quarta aula do curso de Vigilância em Saúde apoiado por plataforma *Business Intelligence* (BI)! Nesse curso estamos aprendendo alguns conceitos de Vigilância Epidemiológica apoiada por ferramentas de BI para análise de dados temporais.

O objetivo do nosso curso é ajudar a análise epidemiológica em um contexto de vigilância em serviços de saúde, utilizando ferramentas que facilitem a vida dos profissionais de saúde.

Essa é nossa primeira aula prática! Para completá-la é importante que você tenha o R e o RStudio instalados em seu computador, conforme instruído na aula 3. Estamos muito animados para que você participe dessa jornada com a gente!

2. Introdução

Ao trabalhar com dados temos que compreender que tipo de informações pode ser extraídas com uma análise temporal. Compreender os dados temporais nos ajuda a entender os momentos entre uma determinada exposição e o período de desenvolvimento da doença e isso é fundamental para realizarmos ações condizentes com o tempo necessário para a prevenção ou tratamento da doença.

Para compreender o que os dados significam e poder comparar sua ocorrência com outros momentos, precisamos trabalhar com indicadores de saúde. A avaliação do modo de vida de populações é um objeto de interesse de governos e instituições e os indicadores de saúde servem para o monitoramento e ações de Saúde Pública, a níveis locais, nacionais e mundiais (1).

Nesta aula, falaremos sobre o cálculo de taxas de incidência. A incidência é um conceito epidemiológico que define a ocorrência de novos casos no tempo. Esse conceito é fundamental e está intimamente ligado às análises de dados temporais. Para compreender esses dados e para calculá-los, usaremos a incidência de dengue para compreender incidência e o cálculo da (2).

3. Indicadores de Saúde – Taxas de Incidência

A incidência é um indicador de saúde referente à frequência de casos novos de uma determinada doença ou agravo originados de uma população sob o risco de adoecimento no início da observação (1). Isso significa que o conceito de incidência é inerente aos dados temporais. Para calcular a incidência o primeiro passo é definir um período de observação. A partir disso, considera-se uma população exposta a determinada condição e acompanha-se a ocorrência de novos casos do desfecho investigado (2).

O número de novos casos é uma medida importante para entender a magnitude do problema, além de ser importante para a gestão de saúde e administração de recursos. Apesar disso, quando queremos comparar

diferentes períodos ou diferentes territórios, precisamos ponderar o número de casos pelo tamanho da população (1,2). Por exemplo, 100 casos novos em uma população de 1.000.000 de habitantes gera uma preocupação menor que 100 casos em uma população de 1.000 habitantes.

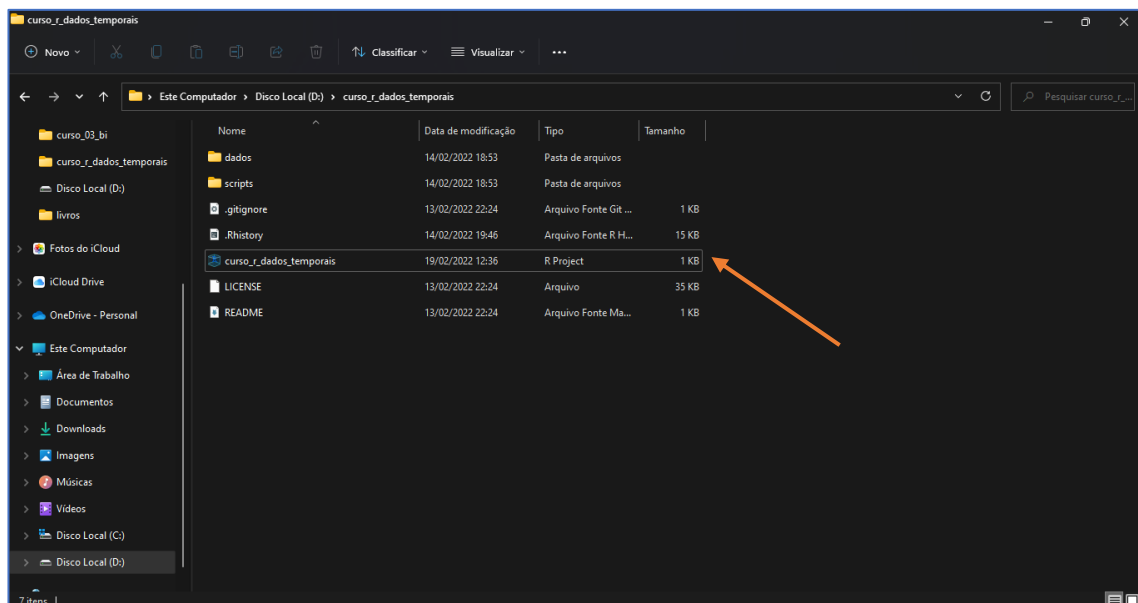
Assim, utilizando o conceito de incidência, temos um indicador de saúde chamado de taxa de incidência, que permite o cálculo do número de novos casos em uma determinada área geográfica e em um determinado período. A fórmula de calcular a taxa de incidência é simples, e é expressa por:

Taxa de Incidência

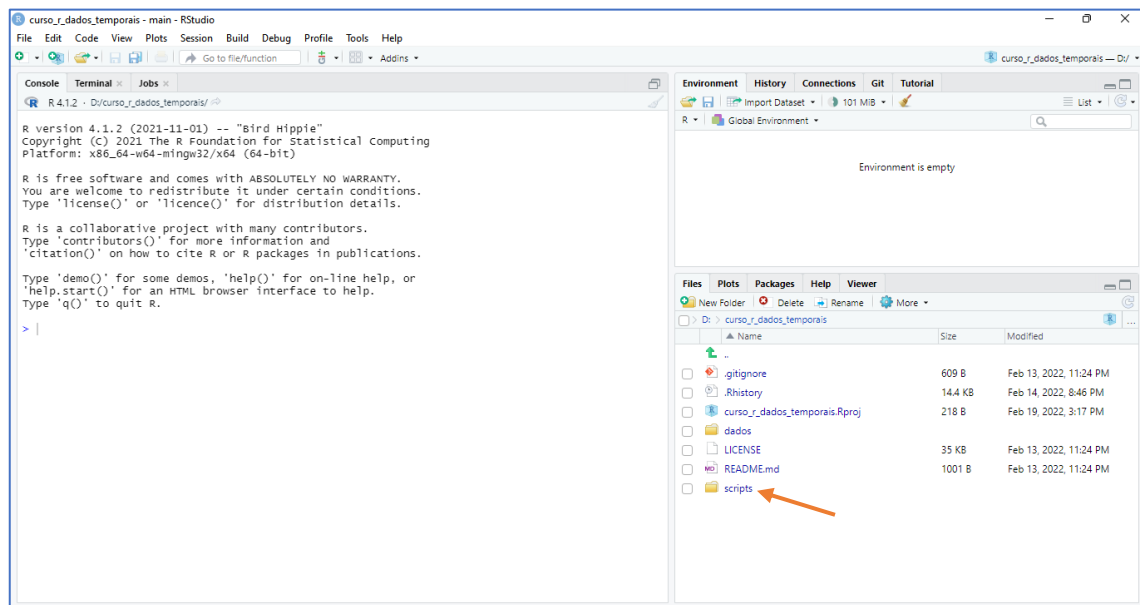
$$= \frac{\text{Número de casos novos em determinado local e período}}{\text{População em determinado período e local}} \times 100.000$$

4. Calculando taxas de incidência com o R

Vamos utilizar o R para calcular taxas de incidência temporais, ou seja, o número de casos em diferentes períodos para o mesmo território. Nosso objetivo é monitorar a taxa de incidência de casos de dengue na população dos estados brasileiros entre 2014 e 2019.



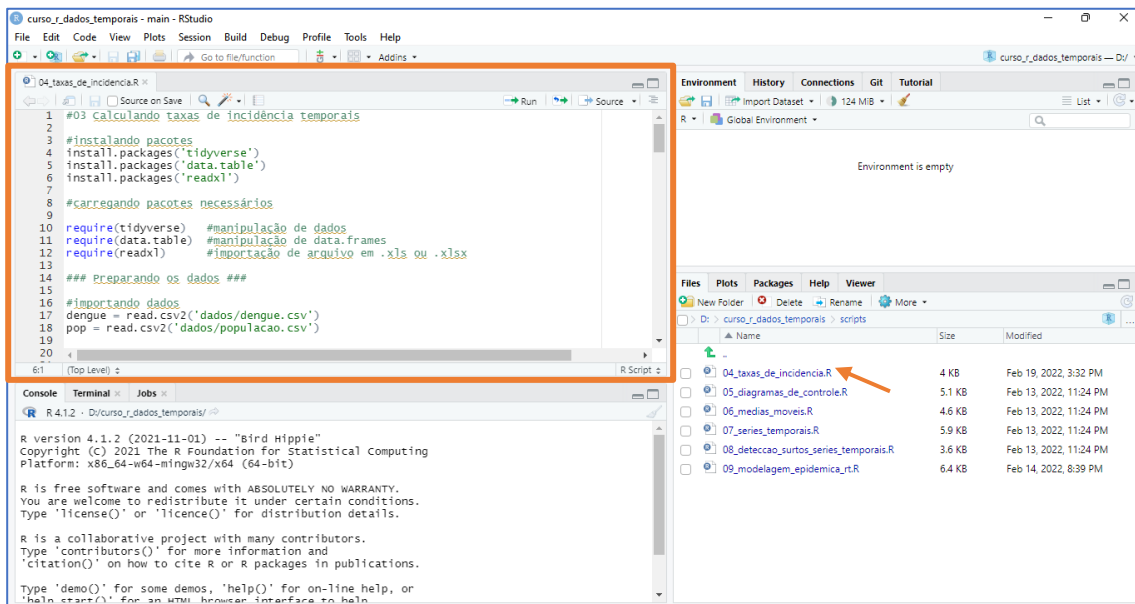
Após abrir o R Project, você receberá uma tela com o *Console* do R (espaço onde o R é executado), o *Environment* (espaço onde ficam dispostos os objetos armazenados pelo R) e a aba *Files* (espaço onde ficam dispostos os arquivos da pasta em que o RStudio está aberto. Você clicará na aba *Files* e escolherá a pasta *scripts* onde estão dispostas todas as aulas desse curso.



Após isso, você escolherá o arquivo “04_taxas_de_incidencia.R” que é referente à aula 4 e o RStudio abrirá o *script* da aula 4. Esse procedimento deverá ser realizado em todas as aulas para a abertura de cada *script* referente a cada uma. Ao abrir o *script*, uma nova janela será aberta, ela é referente ao script e contém todos os códigos que serão executados nessa aula.

O script possui todos os pacotes que serão instalados e carregados nessa aula. Nós vamos usar os pacotes *tidyverse* (para a manipulação e tratamento de dados), o pacote *data.table* (para importação de grandes volumes de dados) e o pacote *readxl* (para a importação de dados de planilhas eletrônicas em formato `.xls` ou `.xlsx`).

Para iniciar o tratamento e importação de dados vamos instalar e carregar os três pacotes para que possamos iniciar os procedimentos de importação e tratamento dos dados.



Para instalar e carregar os pacotes, utilize os seguintes comandos:

```
#instalando pacotes
install.packages('tidyverse')
install.packages('data.table')
install.packages('readxl')

#carregando pacotes necessários
require(tidyverse) #manipulação de dados
require(data.table) #manipulação de data.frames
require(readxl) #importação de arquivo em .xls ou .xlsx
```

Após a instalação e carregamento, vamos importar os dados que estão disponíveis na pasta “dados” dentro do projeto disponibilizado. Não é necessário direcionar o R para a pasta, após abrir utilizando o R Project, os dados já estarão no diretório em que o *script* acessará os dados.

Vamos utilizar dados do *TabNet* com o número de casos de dengue no Brasil entre 2014 e 2019. Além disso, vamos importar também os dados populacionais de cada um dos estados. Esses dados permitirão o

cálculo da taxa de incidência em cada um dos anos considerados no nosso curso.

Para isso, utilizaremos os seguintes comandos:

```
### Preparando os dados ###  
#importando dados  
dengue = read.csv2('dados/dengue.csv')  
pop = read.csv2('dados/populacao.csv')  
  
#Visualizando data.frames  
View(dengue)  
View(pop)  
  
#Verificando nome das colunas  
colnames(dengue)  
colnames(pop)  
  
#modificando nome das colunas de cada data.frame  
colnames(dengue) = c("UF", "2014", "2015", "2016", "2017",  
"2018", "2019", "2020", "2021", "Total")  
colnames(pop) = c("UF", "2014", "2015", "2016", "2017",  
"2018", "2019")  
  
#excluindo colunas que não serão usadas na análise  
dengue$Total = NULL  
dengue$`2021` = NULL  
dengue$`2020` = NULL
```

Para preparar os dados, vamos importar os dados de dengue e os dados populacionais. Após isso, visualizaremos os dados utilizando a função *View*. Os dados de dengue, possuem valores até 2021 e os dados populacionais tem observações até 2019.

Utilizamos a função *colnames* para identificar e mudar o nome das colunas de cada um dos data.frames utilizados. Após isso, vamos excluir

os dados de soma Total, 2020 e 2021 do *data.frame* de dengue. Após isso, vamos realizar a união desses dois *data.frames* e vamos incluir dados de região, para que possamos calcular as taxas de incidência por UF e por região do país.

Utilizando os seguintes comandos, nós vamos criar uma coluna com o código geográfico cadastrado no IBGE para cada estado. Depois disso, vamos importar uma tabela com os códigos geográficos, as regiões e os estados. Esses dados serão úteis para o agrupamento de casos e de populações.

Verificamos a classe dos dados, pois para realizar o relacionamento entre dois *data.frames* é necessário que entender qual o tipo dos dados que serão relacionados. Transformaremos todos os dados de código geográfico em *factor*, usando a função *as.factor*.

Após a transformação dos dados, vamos relacionar os dados usando a função *merge*, usando o código geográfico dos estados. Após esse relacionamento, o *data.frame* dengue possuirá as variáveis de nome do estado e da região.

```
#separando dados de UF
dengue$UF = as.factor(substring(dengue$UF, 1, 2))
pop$UF = as.factor(substring(pop$UF, 1, 2))

#juntando com tabela de estados e regiões
regioes = read_excel("dados/regioes.xlsx")

##ajustando os dados para união das tabelas
class(regioes$COD)
class(dengue$UF)

regioes$COD = as.factor(regioes$COD)

##unindo data.frames
dengue = merge(regioes, dengue, by.x = "COD", by.y = "UF")
```


Para o cálculo da taxa de incidência, precisamos reorganizar o nosso *data.frame*. Quando importamos os dados, nós tínhamos uma tabela com formato *wide*. Isso significa que tínhamos observações (estados) nas linhas e variáveis (anos) nas colunas. Nós vamos transformar as colunas que possuem os dados por ano em duas colunas no formato *long*, nos *data.frames* de dengue e de população. Isso significa que teremos uma coluna com o ano e outra com o número de casos. Após isso, vamos renomear as colunas para que possamos unir os *data.frames* e, após isso, calcular a taxa de incidência. Use os seguintes códigos:

```
#reajustando dados do formato wide to long

##ajustando variaveis
dengue$REGIAO = as.factor(dengue$REGIAO)
dengue$UF = as.factor(dengue$UF)

##mudando o formato dos data.frames
dengue2 = melt(setDT(dengue), id.vars = c("COD","REGIAO",
"UF"), variable.name = "ANO")
pop2 = melt(setDT(pop), id.vars = "UF", variable.name =
"ANO")

##ajustando o nome da variável de valores
names(dengue2)[names(dengue2) == 'value'] <- 'CASOS'
names(pop2)[names(pop2) == 'UF'] <- 'COD'
names(pop2)[names(pop2) == 'value'] <- 'POPULACAO'

#unindo data.frames
dados = dengue2 %>%
  left_join(pop2, by = c("COD", "ANO"))
```

Depois de juntar os *data.frames*, vamos preparar o cálculo da taxa de incidência. Primeiro, vamos criar dois *data.frames* para os dados de casos e população por UF e casos e população por região geográfica. Usando a função *aggregate*, iremos somar o número de casos e de

população por região geográfica. Depois disso, usaremos a função *round* para arredondar a taxa de incidência por UF e por região.

```
#calculando taxa de incidência

#por uf

uf = dados

uf$TX_INC = round(uf$CASOS/uf$POPULACAO*100000, 2)

#por regioao

##agregando os dados por regioao

regiao = aggregate(cbind(CASOS, POPULACAO) ~ REGIAO + ANO,
  data = uf, sum)

##calculando taxa de incidencia por regioao

regiao$TX_INC = round(regiao$CASOS/regiao$POPULACAO*100000,
  2)

#de todo o país

##agregando os dados por regioao

brasil = aggregate(cbind(CASOS, POPULACAO) ~ ANO, data = uf,
  sum)

##calculando taxa de incidencia por regioao

brasil$TX_INC = round(brasil$CASOS/brasil$POPULACAO*100000,
  2)

#criando visualizacoes

##ajustado variável

brasil$ANO = as.double(as.character(brasil$ANO))
```

Após isso vamos criar alguns gráficos de linha e de área para representar a evolução temporal dos dados. Primeiramente, criaremos um gráfico de linhas para a taxa de incidência no Brasil.

```
##gerando gráfico de linhas

ggplot(brasil, aes(x = ANO, y = TX_INC)) +

  geom_line(color = "blue", size = 1) +

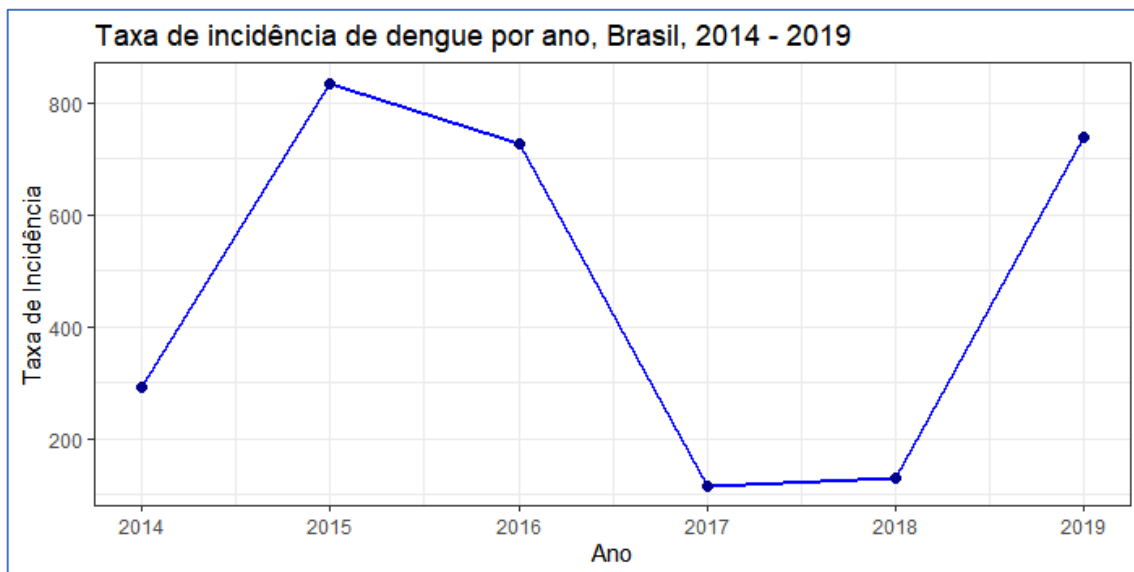
  geom_point(color = "darkblue", size = 2) +

  xlab("Ano") + ylab("Taxa de Incidência") +

  ggtitle("Taxa de incidência de dengue por ano, Brasil,
          2014 - 2019") +

  theme_bw()
```

O gráfico sairá assim:



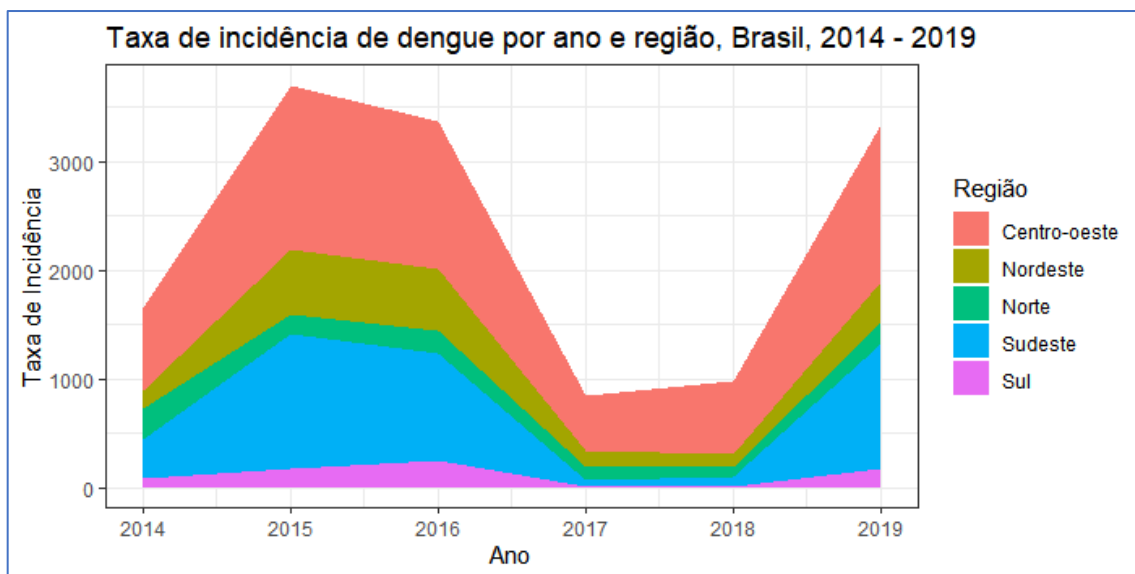
Nós vamos criar um gráfico por região, para podermos comparar as taxas de incidência de cada região.

```
##ajustado variável
regiao$ANO = as.double(as.character(regiao$ANO))

##gerando gráfico de área
ggplot(regiao, aes(x = ANO, y = TX_INC, fill = REGIAO)) +
```

```
geom_area()+
  xlab("Ano") + ylab("Taxa de Incidência")+
  ggtitle("Taxa de incidência de dengue por ano e região,
Brasil, 2014 - 2019")+
  guides(fill=guide_legend(title="Região"))+
  theme_bw()
```

Para esse gráfico, plotamos em formato de área, filtrando cada região por cor.

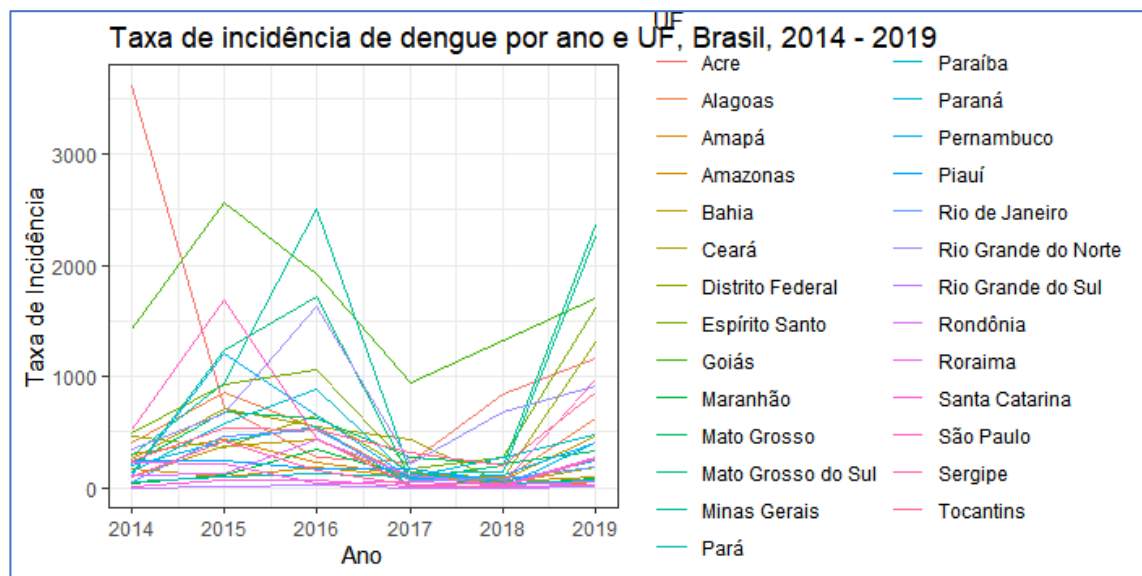


Vamos realizar a plotagem de um gráfico de linhas por UF. Para essa tarefa, usaremos o seguinte código:

```
##ajustado variável
uf$ANO = as.double(as.character(uf$ANO))

##gerando gráfico de múltiplas linhas
ggplot(uf, aes(x = ANO, y = TX_INC, color = UF))+
  geom_line()+
  xlab("Ano") + ylab("Taxa de Incidência")+
  ggtitle("Taxa de incidência de dengue por ano e UF, Brasil,
2014 - 2019")+
  guides(fill=guide_legend(title="Região"))+
  theme_bw()
```

Assim, temos o seguinte gráfico. Note que pelo excesso de variáveis, o gráfico fica ilegível.



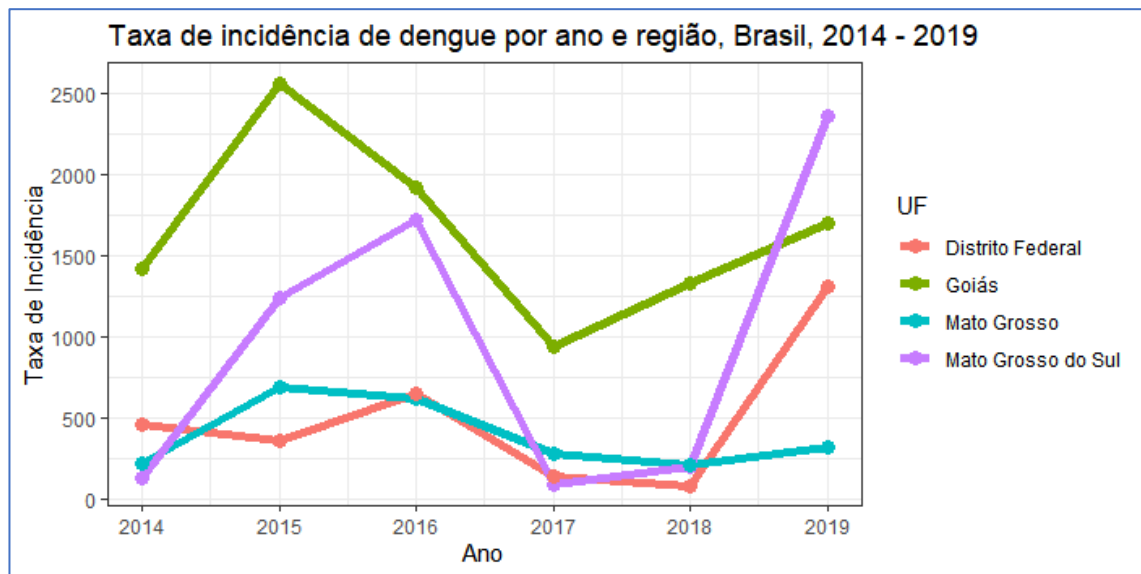
Assim, um gráfico com todas as linhas pode não ser útil para o monitoramento epidemiológico. Nesse sentido, podemos aderir a algumas estratégias que nos ajudem a obter visualizações melhores. Nesse sentido, vamos filtrar os dados por região e vamos criar um gráfico dos estados de cada região.

```
##filtrando por regioao
uf_norte = subset(uf, uf$REGIAO == "Centro-oeste")

##gerando gráfico de múltiplas linhas
ggplot(uf_norte, aes(x = ANO, y = TX_INC, color = UF))+
  geom_line(size = 2)+
  geom_point(size = 3)+
  xlab("Ano") + ylab("Taxa de Incidência")+
  ggtitle("Taxa de incidência de dengue por ano e região,
Brasil, 2014 - 2019")+
  guides(fill=guide_legend(title="Região"))+
  theme_bw()
```

Filtrando pela região Centro-oeste, obteremos o seguinte gráfico.

Para gerar o gráfico das outras regiões, basta trocar o argumento da função para a região desejada.



Referências Bibliográficas

1. Medronho, R. D. A., Bloch, K. V., Luiz, R. R., & Werneck, G. L. (2009). **Epidemiologia**. Editora Atheneu.
2. ROTHMAN, Kenneth; GREENLAND, Sander; LASH, Timothy. **Epidemiologia Moderna-3ª Edição**. Artmed Editora, 2016.

Atividades

1		V	
		F	X
2		V	X
		F	
3		V	X
		F	
4		V	
		F	X
5		V	X
		F	
6		V	
		F	X
7		V	
		F	X
8		V	
		F	X
9		V	X
		F	
10		V	X
		F	
11		V	
		F	X
12		V	X
		F	
13		V	X
		F	
14		V	X
		F	
15		V	
		F	X