

# **Curso de Vigilância em Saúde apoiado por plataforma BI**

## **Aula 05- Criando diagramas de controle**

## **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.

Em nossa aula, vamos criar um diagrama de controle. Os diagramas de controle são ferramentas simples e potentes na identificação de surtos e de epidemias. Vamos utilizar médias históricas e limites inferiores e superiores esperados para as semanas epidemiológicas para estimar taxas esperadas de incidência de dengue.

## **2. Introdução**

Quando falamos de surtos e epidemias, sempre utilizamos uma definição de valores esperados. Mas o que são os valores esperados de uma determinada doença? Qual é o limite de casos que pode ser considerado normal e quando temos valores inesperados? Para responder essas questões, precisamos de dados temporais e de um conhecimento básico estatístico.

Em geral, o conceito de valores esperados está associado às variáveis categóricas e são representados por uma média ponderada em somamos a multiplicação da frequência de cada categoria pela probabilidade de sua ocorrência. Assim, temos uma expectativa da ocorrência daquela categoria no contexto em que ele foi estimado (1).

No caso das taxas de incidência, temos uma variável numérica contínua, então o conceito de valor esperado aplicado a esse contexto não é o mesmo aplicado aos dados categóricos. Entretanto, estudar a ocorrência histórica pode nos indicar quais são as taxas esperadas para cada período e isso pode ser utilizado para o monitoramento de agravos e a identificação de surtos e epidemias (2).

## **3. Diagramas de controle**

Existem diversas técnicas que podem ser aplicadas para a identificação de surtos ou epidemias. O diagrama de controle é uma das técnicas mais simples para essa finalidade. Em geral, essa técnica considera o número de casos no mesmo período anual considerando anos anteriores (2).

O diagrama de controle é um gráfico em que a incidência da doença é representada no eixo das ordenadas (eixo Y) e o período é apresentado no eixo das abcissas (eixo X). Podemos calcular os limites esperados considerando períodos de semana epidemiológica (SE) ou meses do ano. Existem várias técnicas para a construção de limites superiores e inferiores. Nessa aula, utilizaremos a média história e a distribuição normal padrão (2).

Para a construção da média histórica realizaremos os seguintes passos, conforme Medronho, 2016 (2):

1. Organizar os dados pelo período desejado. Em geral, utiliza-se o período de 10 anos anteriores. Em nossa aula, usaremos a semana epidemiológica entre 2015 e 2019 no Distrito Federal.
2. Calcular as taxas de incidência por cada período desejado, isto é, calcular a taxa de incidência para cada uma das semanas entre 2015 e 2018. Sabemos que houve uma epidemia no DF em 2019 e usaremos os dados desse ano para validar nosso diagrama de controle. Nesse caso, calcularemos a taxa de incidência semanal, utilizando a população estimada para o ano.
3. Calcular o desvio padrão para cada SE, com a finalidade de conhecer a dispersão dos valores em torno da média em cada período.
4. Calcula-se o intervalo esperado para o período utilizando os valores da distribuição normal padrão. Para uma confiança de 95%, multiplicamos o desvio padrão por 1,96 para o limite superior e por -1,96 para o limite inferior.
5. Plotamos o gráfico com os limites superiores, inferiores e com a média histórica.
6. Para monitorar os dados atuais, plotamos periodicamente (no período utilizado no diagrama de controle) as taxas daquele período.
7. Caso as taxas sejam superiores ao limite superior esperado para o período, é possível que um possível surto ou epidemia esteja se iniciando. Medidas de controle são esperadas nesse período.

#### **4. Criando um diagrama de controle com o R e RStudio**

Inicialmente, vamos abrir o R Project e, depois disso, o script "05\_diagramas\_de\_controle.R". Todo o processo é descrito de forma detalhada na aula 4. Caso tenha dúvidas, consulte essa aula.

Lembre-se, para acompanhar a aula com qualidade, é necessário abrir o arquivo R Project, ele permite que direcionemos a pasta em que

o R está sendo executado, permitindo a importação correta dos dados. Você poderá baixar o arquivo em nossa plataforma.

Após abrir o arquivo, vamos instalar e carregar os pacotes necessários para a nossa aula.

```
#instalando pacotes
install.packages('tidyverse')
install.packages('data.table')
install.packages('read.dbc')
install.packages('lubridate')

#carregando pacotes necessários
require(tidyverse)    # manipulação de dados
require(data.table)   # manipulação de data.frames
require(read.dbc)     # importação de arquivo em .dbc
require(lubridate)    # tratamento de datas
```

Após isso, vamos importar os dados de população e vamos filtrar o Distrito Federal, para o qual o código geográfico é de número 53.

```
#tratando e selecionando dados de população
pop = read.csv2('dados/populacao.csv')

colnames(pop) = c("UF", "2014", "2015", "2016", "2017",
"2018", "2019")

pop$UF = as.factor(substring(pop$UF, 1, 2))

pop2 = melt(setDT(pop), id.vars = "UF", variable.name =
"ANO")

names(pop2)[names(pop2) == 'UF'] <- 'COD'

names(pop2)[names(pop2) == 'value'] <- 'POPULACAO'

pop_df = subset(pop2, pop2$COD == "53")
```

Nós vamos realizar o procedimento de importação dos dados de dengue. Como os dados estão em diversos arquivos diferentes, usaremos o comando *for* para criar uma estrutura de repetição que irá importar todos os arquivos e reunirá todos os dados em um único *data.frame*.

Vamos usar o comando *glimpse* para visualizar as variáveis importadas para o *data.frame* de dengue.

```
#importando dados de dengue
files      =      cbind(paste0("dados/sinan_dengue_df/",
list.files("dados/sinan_dengue_df/", pattern = "\\*.dbc$")))
d_fn = NULL
for (i in 1:length(files)) {

  d_temp = read.dbc(files[i])
  d_fn= data.frame(rbind(d_fn, d_temp))
}
glimpse(d_fn)
```

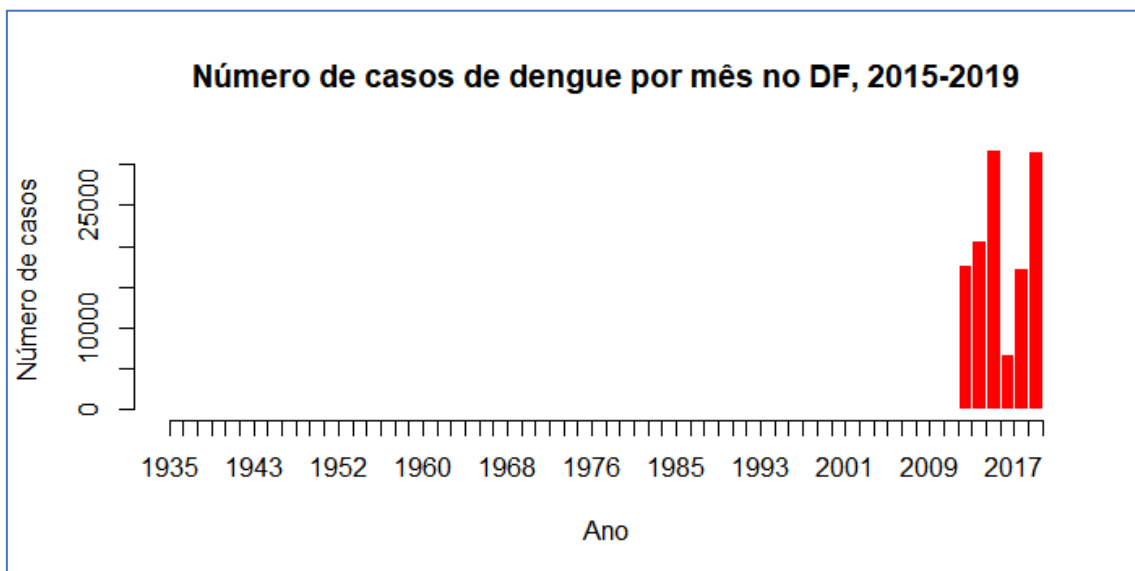
Verificaremos se a variável de data dos primeiros sintomas está classificada como um objeto da classe *date*. Isso é importante para que o R compreenda que estamos lidando com dados temporais. Sempre devemos usar a data dos primeiros sintomas para o monitoramento de doenças infecciosas. Essa data permite estimar a data de infecção e é fundamental para a gestão de medidas de controle de doenças.

```
#analizando o tipo da variável de data
class(d_fn$DT_SIN_PRI)

#analizando a amplitude dos dados
summary(d_fn$DT_SIN_PRI)
hist(d_fn$DT_SIN_PRI, breaks = 60,
      main = "Número de casos de dengue por mês no DF, 2015-2019",
      xlab = "Ano", ylab = "Número de casos",
      freq = T,
      border = "White", col = "red")
```

```
#ajustando a variável data
d_fn$DT_SIN_PRI = as.Date(d_fn$DT_SIN_PRI, "%Y-%m-%d")
d_fn = subset(d_fn, d_fn$DT_SIN_PRI > "2014-12-31" &
d_fn$CLASSI_FIN != "5")
```

Investigaremos se as datas apresentam a amplitude esperada. É possível que existam datas erradas, por exemplo um erro de digitação pode fazer com que a data dos primeiros sintomas seja colocada como 01 de janeiro de 1900 ou 31 de dezembro de 2067. Para essas datas impossíveis e discrepantes, podemos excluir as observações para que elas não influenciem de forma negativa a nossa análise. A análise da amplitude será realizada por meio de um histograma.

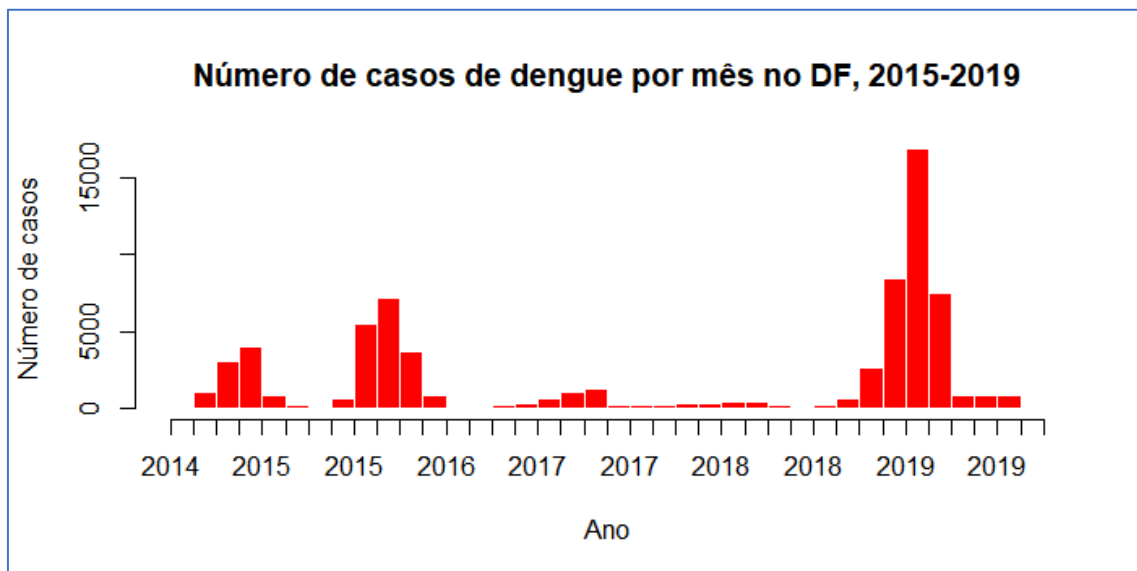


Observe que o histograma mostra datas para o ano de 1935. Essas datas não interessam à nossa análise e, provavelmente, devem ter sido cadastradas de forma equivocada. Por isso, vamos filtrar apenas os dados de datas superiores a 01 de janeiro de 2014 e que possuem classificação final diferente de 05, que se refere aos casos descartados. Consideraremos apenas casos confirmados e casos prováveis.

```
#ajustando a variável data
d_fn$DT_SIN_PRI = as.Date(d_fn$DT_SIN_PRI, "%Y-%m-%d")
d_fn = subset(d_fn, d_fn$DT_SIN_PRI > "2014-12-31" &
d_fn$CLASSI_FIN != "5")
```

```
#analizando a amplitude dos dados após o corte
summary(d_fn$DT_SIN_PRI)
hist(d_fn$DT_SIN_PRI, breaks = 60,
     main = "Número de casos de dengue por mês no DF, 2015-2019",
     xlab = "Ano", ylab = "Número de casos",
     freq = T,
     border = "White", col = "red")
```

Após isso, vamos plotar um novo histograma para verificar como os dados ficaram após o filtro.



Com as datas ajustadas, percebemos que ao analisar o histograma, a dengue apresenta um comportamento sazonal, que aumenta em determinados períodos. Apesar disso, os anos de 2017 e 2018 tem incidência menor que 2014 e 2015. Em compensação, o ano de 2019 apresenta uma incidência bem maior que os anos anteriores.

Vamos calcular a taxa de incidência de dengue para o período considerado e vamos plotar um gráfico das taxas por semana epidemiológica.

```
#criando tabela de casos por semana epidemiológica
d_fn$CASO = 1
```



```
d_week = aggregate(CASO ~ SEM_PRI, data = d_fn, sum)

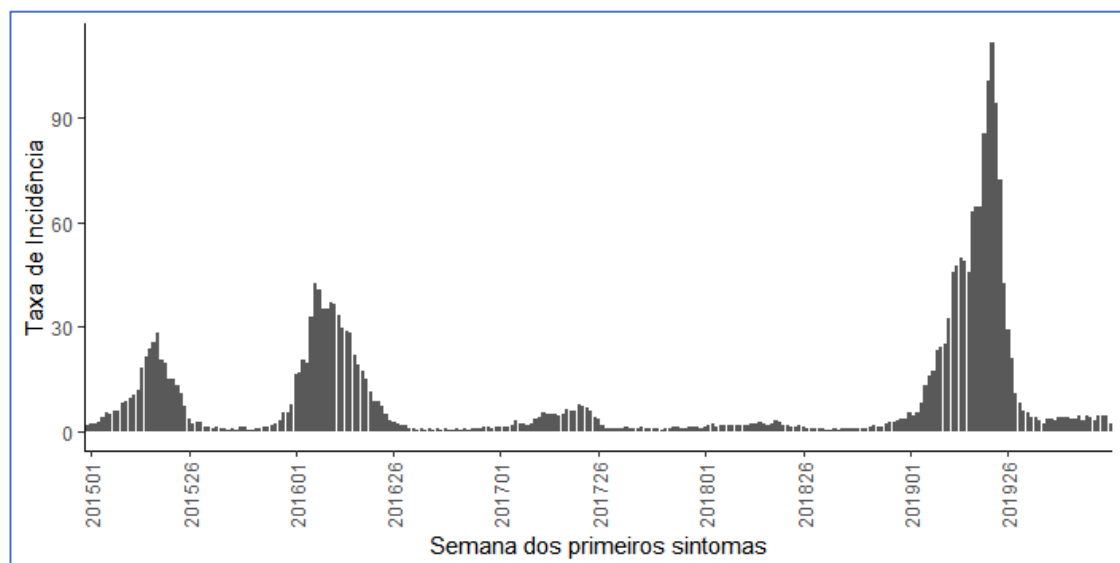
#unindo dados de população
d_week$ANO = substring(d_week$SEM_PRI, 1, 4)
d_week = merge(d_week, pop_df, by = "ANO")

#calculando taxa de incidência por semana epidemiológica
d_week$TX_INC = round(d_week$CASO/d_week$POPULACAO*100000,
  2)

ggplot(d_week, aes(x = SEM_PRI, y = TX_INC))+
  geom_bar(stat = 'identity')+
  xlab("Semana dos primeiros sintomas")+
  ylab("Taxa de Incidência")+

  scale_x_discrete(breaks=c("201501","201526","201601",""
    201626",    "201701","201726",    "201801",    "201826",
    "201901","201926"))+
  theme_classic()+
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5,
    hjust=1))
```

O gráfico apresenta muita semelhança com o gráfico anterior.



Ainda estamos trabalhando com os dados de forma longitudinal, ou seja, considerando as taxas de incidência em cada período de forma separada.

Com os dados preparados, calcularemos a média de casos por semana epidemiológica entre 2014 e 2018. Também calcularemos o desvio padrão e os limites superiores e inferiores para esses anos.

Vamos separar os dados de 2019 para incluir como ano de interesse em nossa análise. O objetivo é verificar se houve epidemia nesse ano, considerando os níveis esperados de dengue para todo o ano. Todos esses dados serão úteis para a plotagem do nosso diagrama de controle.

```
#calculando média semanal da taxa de incidência
d_week2 = data.frame(ANO = d_week$ANO,
                      SEM_PRI = d_week$SEM_PRI,
                      TX_INC = d_week$TX_INC)
d_week2$SEM_PRI = substring(d_week2$SEM_PRI, 5, 6)

d_week2 = reshape2::dcast(d_week2, SEM_PRI ~ ANO, value.var
= "TX_INC")

#removendo ano desconsiderado na análise
d_week2$`2014` = NULL

#gerando data frame para o diagrama de controle
d_week3 = d_week2
d_week3$`2019` = NULL

#gerando data.frame para o teste do controle
d_week4= data.frame(X2019 = d_week2$`2019`)

#gerando estatísticas
row.names(d_week3) = d_week3$SEM_PRI
row.names(d_week4) = d_week3$SEM_PRI
d_week3$SEM_PRI = NULL
```

```

d_week3 = d_week3[1:52,1:4]
d_week4 = data.frame(X2019 = d_week4[1:52,1])

d_week3$mean = round(rowMeans(d_week3), 2)
d_week3$sd = round(apply(d_week3[,5], 1, sd), 2)

#criando os limites superiores e inferiores do diagrama
d_week3$low = round((d_week3$mean - 1.96*(d_week3$sd)), 2)
d_week3$high = round((d_week3$mean + 1.96*(d_week3$sd)), 2)

d_week3$low = ifelse(d_week3$low < 0, 0, d_week3$low)

d_week3$week = 1:52
d_week4$week = 1:52

```

Após isso, vamos plotar nosso diagrama de controle, considerando a média histórica e os limites inferiores e superiores entre os anos de 2014 e 2018. Utilizamos as taxas de incidência de 2019 para compreender se houve epidemia nesse ano.

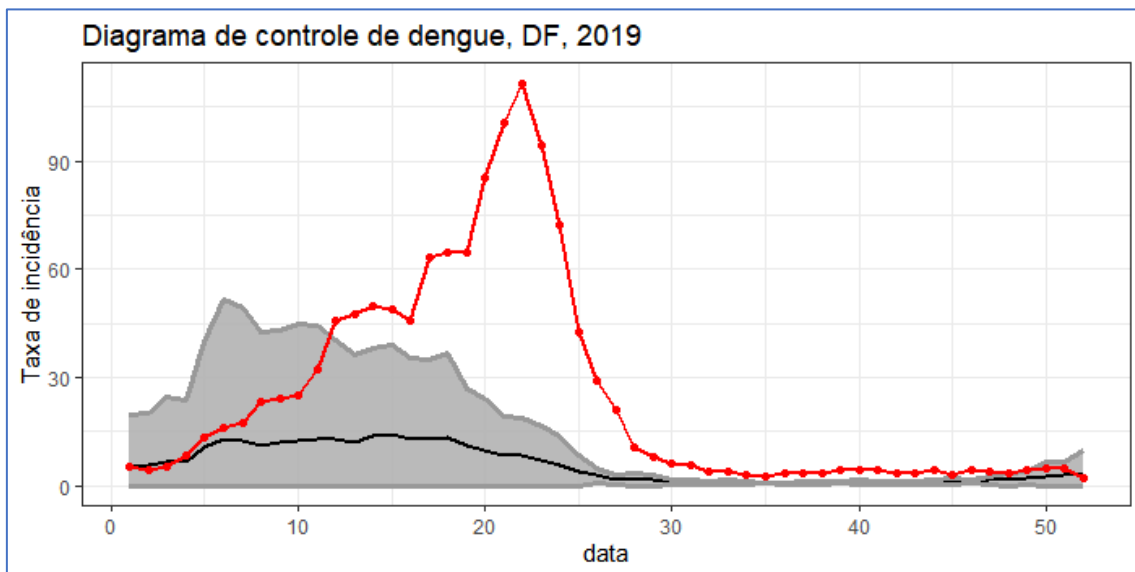
```

ggplot()+
  geom_ribbon(data = d_week3, aes(ymin=low, ymax=high, x =
week), alpha=0.9 , fill = "grey70")+
  geom_line(data = d_week3, aes(x = week, y= mean), colour
= "black", size = 1)+
  ggtitle("Diagrama de controle de dengue, DF, 2019")+
  theme_bw()+
  ylab("Taxa de incidência")+
  xlab("data")+
  geom_line(data = d_week3, aes(x = week, y= low), colour =
"grey60", size = 1.3)+
  geom_line(data = d_week3, aes(x = week, y= high), colour
= "grey60", size = 1.3)+

```

```
geom_line(data = d_week4, aes(x = week, y= X2019), colour
= "red", size = 1)+
geom_point(data = d_week4, aes(x = week, y= X2019), colour
= "red", size = 1.5)
```

O gráfico mostra que após a 10ª semana de 2019, a taxa de incidência foi muito maior do que o esperado historicamente.



Em uma situação de monitoramento, podemos plotar as taxas semana após semana para realizar medidas de controle quando as primeiras taxas superiores ao limite foram encontradas. Nesse sentido, simulamos como ficaria o diagrama de controle na 13ª SE, quando as primeiras taxas superiores foram encontradas.

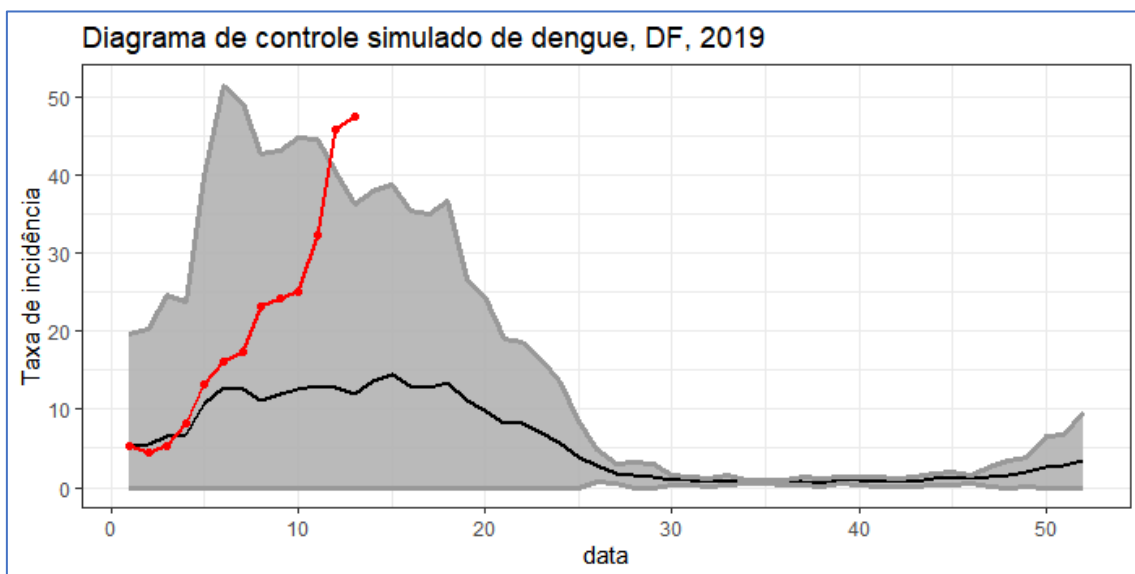
```
ggplot()+
  geom_ribbon(data = d_week3, aes(ymin=low, ymax=high, x =
week), alpha=0.9 , fill = "grey70")+
  geom_line(data = d_week3, aes(x = week, y= mean), colour
= "black", size = 1)+
  ggtitle("Diagrama de controle de dengue, DF, 2019")+
  theme_bw()+
  ylab("Taxa de incidência")+
  xlab("data")+
```

```

geom_line(data = d_week3, aes(x = week, y= low), colour =
"grey60", size = 1.3)+
  geom_line(data = d_week3, aes(x = week, y= high), colour
= "grey60", size = 1.3)+
  geom_line(data = d_week4, aes(x = week, y= X2019), colour
= "red", size = 1)+
  geom_point(data = d_week4, aes(x = week, y= X2019), colour
= "red", size = 1.5)

```

O gráfico, disposto a seguir, poderia indicar à Vigilância em Saúde que medidas de controle de insetos e educação popular deveriam ser reforçados na população do Distrito Federal. A ausência de medidas efetivas, elevou a taxa de incidência, que foi maior que o esperado até o final daquele ano.



## Referências Bibliográficas

1. BRUCE, Andrew; BRUCE, Peter. **Estatística Prática para Cientistas de Dados**. Alta Books, 2019.
2. Medronho, R. D. A., Bloch, K. V., Luiz, R. R., & Werneck, G. L. (2009). **Epidemiologia**. Editora Atheneu.
3. ROTHMAN, Kenneth; GREENLAND, Sander; LASH, Timothy. **Epidemiologia Moderna-3ª Edição**. Artmed Editora, 2016.

## Atividades

|    |  |   |   |
|----|--|---|---|
| 1  | Epidemias se relacionam a um número sempre constante de casos de doenças em diferentes territórios   | V |   |
|    |  | F | X |
| 2  | Quando falamos de surtos e epidemias nos referimos a um aumento inesperado no número de casos de um determinado agravo em determinado território                         | V | X |
|    |  | F |   |
| 3  | Taxas de incidência são calculadas com o número de casos ponderado pelo número de indivíduos no território   | V | X |
|    |  | F |   |
| 4  | Taxas de incidência resultam em variáveis categóricas  | V |   |
|    |  | F | X |
| 5  | Existem diversas técnicas que podem ser aplicadas para a identificação de surtos ou epidemias.   | V | X |
|    |  | F |   |
| 6  | O diagrama de controle é uma das técnicas mais simples para monitoramento de surtos e epidemias.   | V |   |
|    |  | F | X |
| 7  | O diagrama de controle não é uma técnica usada para monitoramento de surtos e epidemias.   | V |   |
|    |  | F | X |
| 8  | Não é necessário organizar os dados para elaborar um diagrama de controle.   | V |   |
|    |  | F | X |
| 9  | Uma das etapas da elaboração do diagrama de controle é calcular a taxa de incidência para cada período considerado.  | V | X |
|    |  | F |   |
| 10 | O diagrama de controle é um gráfico em que a incidência da doença é representada no eixo das ordenadas (eixo Y) e o período é apresentado no eixo das abcissas (eixo X). | V | X |
|    |  | F |   |
| 11 | O diagrama de controle é um gráfico de dispersão de pontos.  | V |   |
|    |  | F | X |
| 12 | Estudar a ocorrência histórica pode nos indicar quais são as taxas esperadas para cada período.  | V | X |
|    |  | F |   |
| 13 | Não podemos utilizar os intervalos esperados em um diagrama de controle  | V | X |
|    |  | F |   |
| 14 | Estudar a ocorrência histórica pode ser utilizado para o monitoramento de agravos e a identificação de surtos e epidemias.   | V | X |
|    |  | F |   |
| 15 | Caso as taxas sejam superiores ao limite superior esperado para o período, é possível que um possível surto ou epidemia esteja se iniciando.                             | V | X |
|    |  | F |   |