

UNIVERSIDADE FEDERAL DO ESTADO DO RIO DE JANEIRO (UNIRIO)
CENTRO DE CIÊNCIAS HUMANAS E SOCIAIS (CCHS)
ESCOLA DE BIBLIOTECONOMIA (EB)

Relatório de Estatística

Incidentes de Segurança da Informação no Brasil

Myllena Cristina da Silva Lúcio – 20191332037- myllena.lucio@edu.unirio.br

Rômulo Silveira de Barros de Oliveira – 20191332023 – romulo.s.oliveira@edu.unirio.br

Stephany França de Melo – 20191332050 – stephany.melo@edu.unirio.br

Sumário

1.Introdução 2.Objetivo 3.Metodologia 4.Análise base de dados 4.1.Dicionário de Dados
4.2.Tamanho da Amostra 4.3.Construção dos Indicadores 5.Análise de Resultados
6.Conclusão

Rio de Janeiro

2021

1.Introdução

Desde que a internet se estabeleceu no mundo, novas formas de cometer crimes foram criadas: golpes, roubos de informação, vírus , entre outros. Os infratores são mais conhecidos como hackers. Para combater esses infratores, os países tiveram que criar medidas que pudessem para-los. No Brasil, houve uma grande demora para a implementação de uma resposta a esses cibercrimes e com isso a tecnologia foi avançando juntamente com os ataques e invasões.

Com esse combate tardio no Brasil, o reflexo na quantidade de crimes e ataques foram tendo maiores proporções, sendo isso o principal assunto apontado na base de dados “Incidentes de Segurança da Informação no Brasil”.

As invasões e os ataques de vírus propagados de forma a prejudicar os outros é um problema real, que assola ao povo brasileiro. Com isso, deixando o país em 6º lugar sendo considerado um dos mais vulneráveis a vírus que sequestram informações de computadores.

2.Objetivo

O trabalho tem como objeto de estudo a base de dados “Incidentes de Segurança da Informação no Brasil” retirada da plataforma Kaggle, com o objetivo principal de averiguar o grau de aumento ou diminuição de invasões e ataques por vírus de computador conforme o avanço da tecnologia no País.

Os objetivos específicos do estudo partem do pressuposto de:

- Analisar o número total de incidentes de *worm* relatos;
- Analisar o ano e mês com maior ou menor ocorrência de invasões;
- Analisar o número total de incidentes de invasão relatados.

3. Metodologia

A base de dados escolhida trata-se dos Incidentes de Segurança da Informação no Brasil, recolhida da plataforma Kaggle. Nela não se consta os dados do ano de 2020, por isso, a base de dados foi atualizada pelo grupo tendo as informações retiradas do site do Centro de Estudos, Resposta e Tratamento de Incidentes de Segurança no Brasil.

A ideia inicial do trabalho era investigar as diferenças e possíveis melhorias nos incidentes de segurança da informação após a Lei Geral de proteção de Dados ter entrado em vigor em setembro de 2020, porém, ainda não se tem dados de 2021. Sendo assim, o grupo não achou que a investigação seria satisfatória, tendo em vista que, temos uma diferença de 3 meses.

A base de dados conta com 10 colunas e 133 linhas. Foi decidido pelo grupo que iríamos trabalhar com duas variáveis qualitativas e duas quantitativas, a fim de comparar a quantidade de invasões e Worms (programas independentes, “vírus” de computador).

Contudo, o grupo se deparou com dois problemas: os anos não estavam contando como variável qualitativa por se tratarem de números, tendo como a solução acrescentar a palavra “Ano” antes de cada número, fazendo com que o programa R reconhecesse os anos como variáveis qualitativas. O segundo problema foi a dificuldade de construir gráficos variados com a base de dados escolhida.

Para que fosse possível uma melhor visualização dos totais de incidentes de segurança por ano, o grupo decidiu elaborar uma nova base de dados a partir da que já tínhamos, somente com os totais.

4. Análise da base de dados

No Boxplot da **figura 1**, é possível observar que no ano de 2020 teve um grande aumento de ataques de vírus, fazendo com o que a mediana do próprio ano seja maior que os *outliers* dos anos de 2012 e 2014, além do ano que se antecede a ele. O grupo com maior dispersão é o ano de 2020.

Figura 1. Elaborado pelos autores.

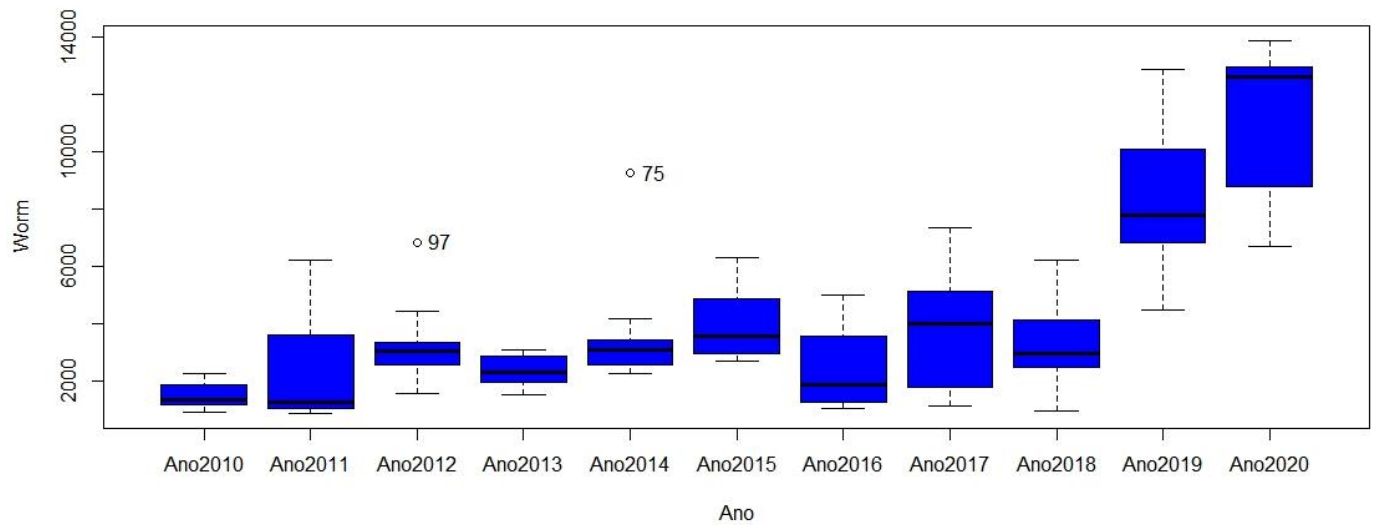
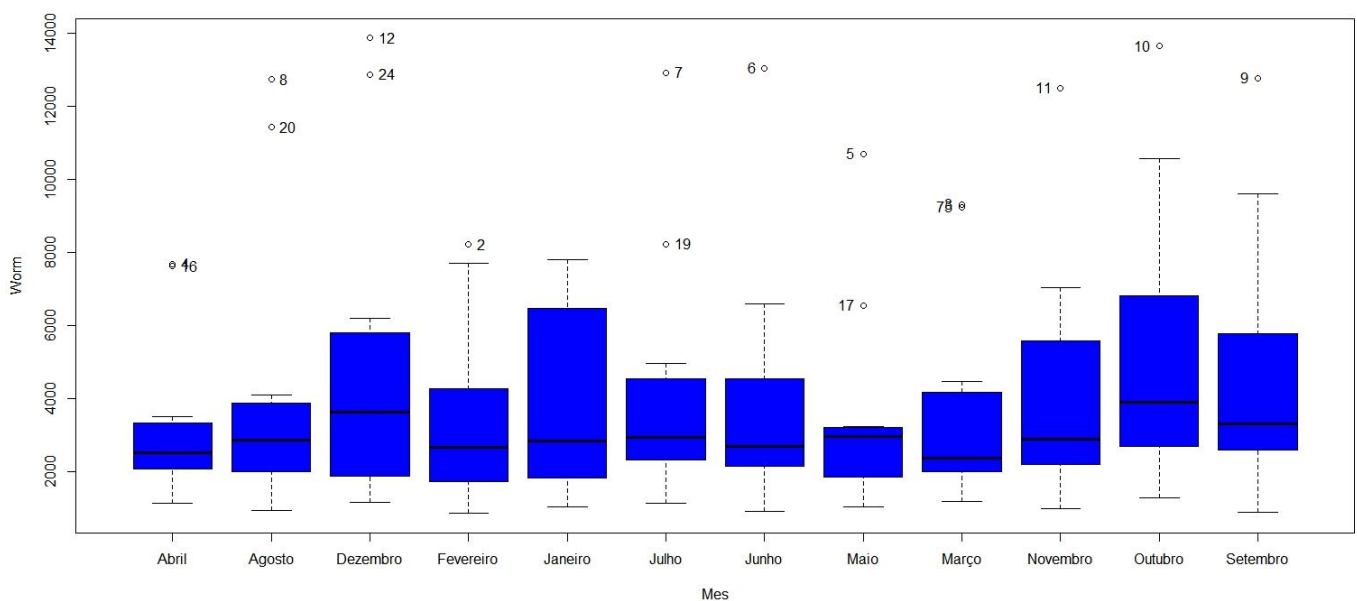


Figura 2. Elaborado pelos autores



No Boxplot acima (**fig.2**), podemos perceber que Março, Abril, Maio, Julho e Agosto são os meses que menos tiveram ocorrências de ataques de vírus. O mês de Dezembro possui sua caixa levemente assimétrica positiva pois sua mediana está mais próxima do primeiro quartil.

Figura 3. Elaborado pelos autores

			Invasão por ano						
Anos	Média	Desvio Padrão	IQR	Mínimo	25%	Mediana	75%	Máxima	Total
Ano2010	7,416667	6,666856	6,75	1	2,75	5	9,5	22	12
Ano2011	8,833333	5,87496	4,5	2	5,75	7	10,25	25	12
Ano2012	651,25	306,480943	193,25	100	532,25	611	725,5	1320	12
Ano2013	933,9167	580,261925	767,25	363	515	666,5	1282,25	2149	12
Ano2014	542,4167	395,450826	466,25	187	270	443	736,25	1559	12
Ano2015	204,75	124,483022	70,25	104	134,75	153,5	205	489	12
Ano2016	141,25	78,003642	35,75	41	103,25	125,5	139	320	12
Ano2017	33,41667	26,654211	24,75	1	17	23,5	41,75	99	12
Ano2018	89,58333	55,896753	77	11	44,75	89,5	121,75	193	12
Ano2019	43,91667	21,360364	29	19	27	41	56	80	12
Ano2020	101	42,177116	56	42	66,75	104	122,75	184	12
Tabela 1. Elabora pelos autores									

A tabela acima (**fig.3**) busca fazer a análise das invasões de redes conectadas à internet no Brasil. Ela evidencia que nos anos de 2012, 2013 e 2014 a média dos crimes de invasão é superior aos demais. Nos outros anos, a mediana aponta os mesmos resultados, pois os valores são bem próximos.

Figura 4. Elaborado pelos autores

			Invasão por mês						
Meses	Média	Desvio Padrão	IQR	Mínimo	25%	Mediana	75%	Máxima	Total
Abril	347,1818	590,5324	144	3	27,5	110	171,5	1559	11
Agosto	165	223,8678	129,5	5	32,5	55	162	659	11
Dezembro	210,9091	263,1446	308	5	26	99	334	727	11
Fevereiro	275,3636	389,2131	339,5	5	31	95	370,5	1206	11
Janeiro	386,1818	628,0702	463	1	36,5	83	499,5	2149	11
Julho	235	387,3425	160,5	1	24,5	106	185	1320	11
Junho	226,0909	252,8436	355	2	33	127	388	619	11
Maiο	218,7273	278,0389	173	5	62	125	235	972	11
Março	347,1818	519,5725	500	1	24	93	524	1686	11
Novembro	191	209,7146	201	6	59,5	120	260,5	716	11
Outubro	206,3636	280,0251	186,5	5	31,5	86	218	754	11
Setembro	199,4545	308,686	191,5	10	26	71	217,5	1043	11

Na **figura 4**, podemos observar quais meses tiveram a média de invasões inferiores aos demais, sendo eles: Julho, Agosto e Novembro.

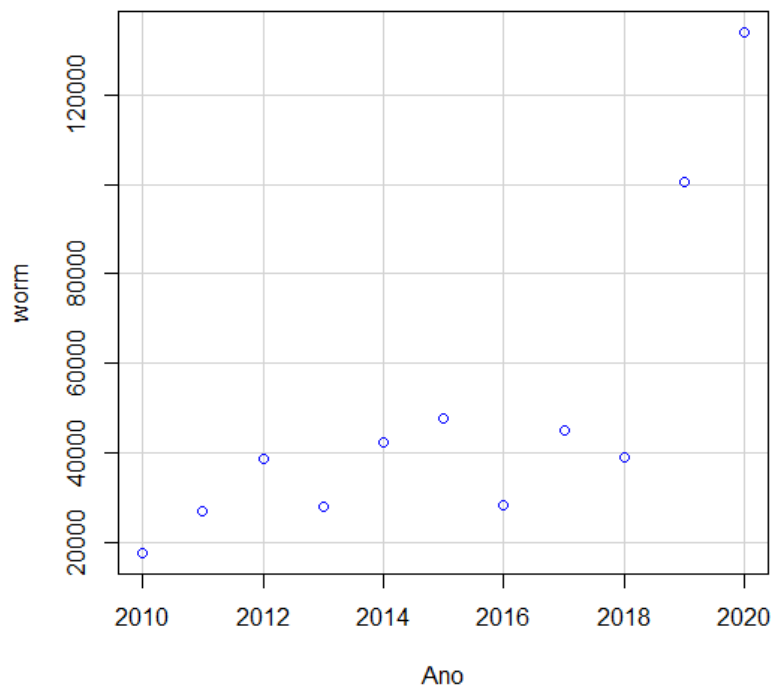


Figura 5.
Elaborado pelos
autores

No gráfico acima (**fig.5**), o *Worm* tem correlação positiva, fraca e teve o seu eixo x composto pelos anos e o seu eixo y pelas ocorrências por ano.

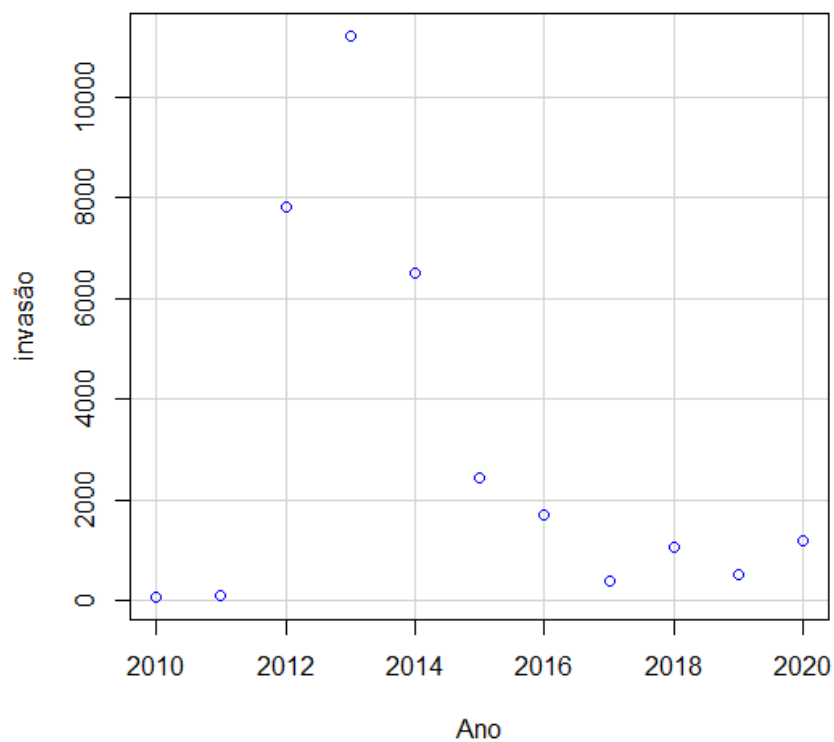


Figura 6.
Elaborado
pelos autores

O gráfico invasão, como podemos ver na **figura 6** acima, não tem correlação. E assim como o gráfico *worm*, seus eixos são 'ano' e 'ocorrência', tendo a mesma ordem. Porém ao contrário do gráfico da **figura 5**, as invasões deste gráfico vem diminuindo conforme o passar do tempo.

4.1.Dicionário de Dados

Worm - Tipo de malware/ Vírus de computador;

DOS - Disk Operating System/ Negação de Serviço no Sistema Operacional em disco;

Scan - Incidentes de varredura.

4.2.Tamanho da Amostra

A base de dados utilizada pelo grupo conta com 10 colunas e 133 linhas.

4.3. Construção dos Indicadores

Taxa de *Worms* = Número de ataques de vírus;

Taxa Mês = Meses de ocorrências de incidentes;

Taxa Ano = Anos com mais ocorrências de incidentes;

Taxa de Invasões = Número de invasões.

5.Análise de Resultados

Mediante à análise dos gráficos, percebe-se que apesar de existir uma discrepância na quantidade de ataques com *Worms* nos anos de 2012 e 2014, os anos que mais tiveram ataques desse tipo de “vírus” de computador foram nos anos de 2019 e 2020.

Contabilizando todos os anos para visualizar em quais meses os ataques mais acontecem, observa-se que existe uma quantidade de *outliers* muito grande em vários meses, contudo, os meses que tendem a ter mais ataques são Janeiro, Setembro e Outubro.

Para fazer a análise das invasões das redes conectadas à Internet no Brasil, utilizamos tabelas de resumo. A tabela revela que nos anos de 2012, 2013 e 2014, a média de invasões foi muito superior que nos outros anos.

Quando analisamos os meses, foi mais interessante observar os que tiveram uma média de invasões bem inferior em relação aos outros, são eles: julho, agosto e novembro.

6.Conclusão

Conclui-se que ataques com vírus perigosos de computador são mais frequentes numa sociedade mais tecnológica a partir de 2017. No entanto, em relação as invasões a sistemas conectados à internet os resultados não compreendem a premissa do estudo, pois, apontam uma grande porcentagem entre 2012 e 2014.

REFERÊNCIA

CERT.br, 2020. Disponível em: <https://www.cert.br/>. Acesso: 20 de nov. 2020

Kaggle, 2019. Disponível em: <https://www.kaggle.com/rodrigoriboldi/incidentes-de-segurana-da-informao-no-brasil>. Acesso 20 de set. 2020