Análise dos concursos da

Mega-Sena

­­

Universidade Federal Rural de Pernambuco

Computação para Análise de Dados – 2018.1

Bruno Henrique Pereira Marques

Danielly de Moura Borba Queiroz

Recife, 12 de Agosto de 2018

# **Introdução**

No Brasil, milhões de pessoas jogam nas loterias federais, seja com baixa frequência ou tem o costume de apostar. Assediadas pelos altos valores das estimativas do prêmio, ainda sim, sabem que a probabilidade de acertar as seis dezenas da Mega-Sena é de 1 em 50.063.860 (cinquenta milhões, sessenta e três mil e oitocentos e sessenta), apostando em apenas um jogo com 6 dezenas que custa R$3,50. É possível reduzir esta proporção para 1 em 10.003 (dez mil e três) porém, esta aposta custará R$17.517,50 [Ref. 1].

O valor arrecadado em cada concurso da Mega-Sena é dividido nas seguintes partes: 54,7% para Programas Sociais do Governo e 45,3% para a Premiação. Ainda há um desconto de 13,8% da parte da premiação para o imposto de renda. Por fim, o resultado desta operação é dividido em: 35% para 6 acertos, 19% para 5 acertos, 19% para 4 acertos, 22% acumulado para o próximo concurso e 5% acumulado para o concurso no final do ano. Apesar da porcentagem final do prêmio ser baixa, o número de apostadores tem aumentado [Ref. 2].

Através dos dados históricos de todos os concursos da maior loteria do Brasil [Ref. 3], realizados desde o primeiro (em 11/03/1996) até o último concurso (número 2000) do ano de 2017, esperamos revelar dados como: crescimento do número de ganhadores, crescimento do valor dos prêmios, estado e região com maior número de ganhadores, estado e região com maior valor premiado, e relações entre a ocorrência das dezenas.

Para explorar os dados da Mega-Sena, será utilizado o arquivo no formato HTM disponibilizado no site oficial da loteria [Ref. 1]. Esta é a única forma que os dados são encontrados. A lista a seguir apresenta as principais variáveis contidas no arquivo.

* Número do concurso
* Data do concurso
* Primeira dezena sorteada
* Segunda dezena sorteada
* Terceira dezena sorteada
* Quarta dezena sorteada
* Quinta dezena sorteada
* Sexta dezena sorteada
* Ganhadores da sena
* Cidade dos ganhadores
* Estado dos ganhadores
* Valor de rateio da sena
* Valor acumulado

A proposta deste trabalho é apresentar gráficos de interesse, representados por ano, por estado, por região, pelos valores dos prêmios e com base nas dezenas sorteadas. Esses gráficos ajudarão a apresentar o crescimento da loteria no decorrer dos anos, como o cenário está montado de acordo com os ganhadores no Brasil e como a ocorrência das dezenas está distribuída. Além disso, será apresentada previsões de estimativas de prêmio, valores acumulados e estimativa para a Mega-Sena da virada.

# **Pacotes requeridos**

A tabela seguinte apresenta os pacotes que são necessários para executar o código sem erros.

|  |  |
| --- | --- |
| Nome do Pacote | Descrição |
| Tidyr | Pacote para organização de dados. |
| Dplyr | Pacote para organização de dados. |
| Maps | Desenha linhas e polígonos conforme um banco de dados de mapas. |
| Maptools | Leitura de ESRI *shapefiles* de mapas. |
| Mapproj | Gerencia mapas, projeções e *grids*. |
| Ggmap | Gerencia mapas baseados em ggplot2, com suporte a Google Maps. |
| Rgdal | Conjunto de operações para projeções. |

# **Preparação dos dados**

Os dados inicialmente foram baixados pelo link [Ref.1] onde se encontram no formato HTML. Cada variável está localizada e uma coluna do código HTML conforme a figura 1.

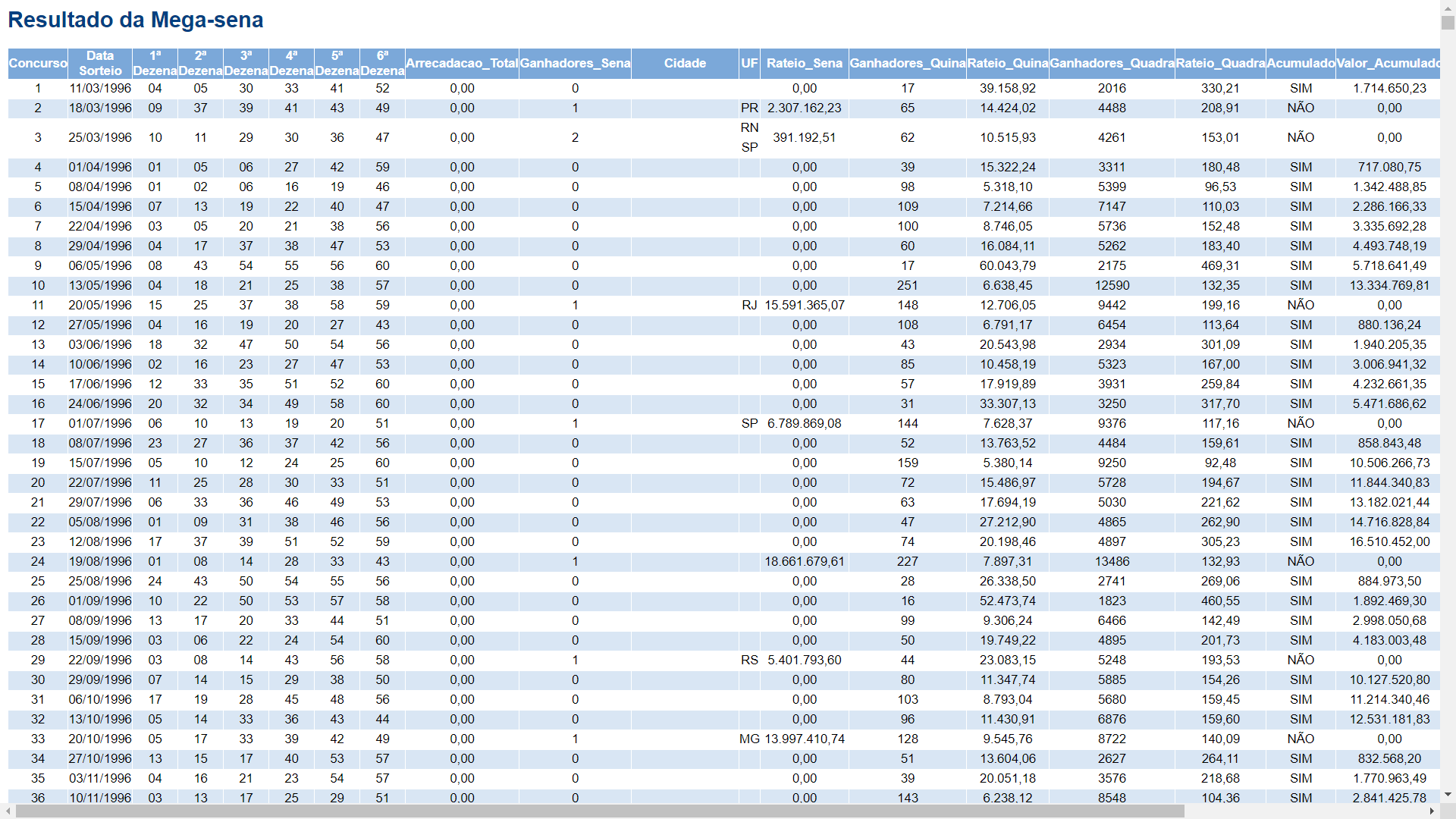


Figura 1: Tabela HTML com os dados da Mega-Sena.

Em seguida, foi realizado uma conversão manualmente do código HTML para CSV utilizando a ferramenta Excel 2013. Dessa forma, passa a ser possível carregar os dados para processamento com a linguagem R. A figura 2 apresenta parte dos dados carregados na ferramenta RStudio. Alguns problemas de conversão são visíveis como a linha 4, por exemplo, a qual está toda preenchia com ***NA***, valores de UF separados por vírgula e a formatação da moeda Rateio\_Sena.

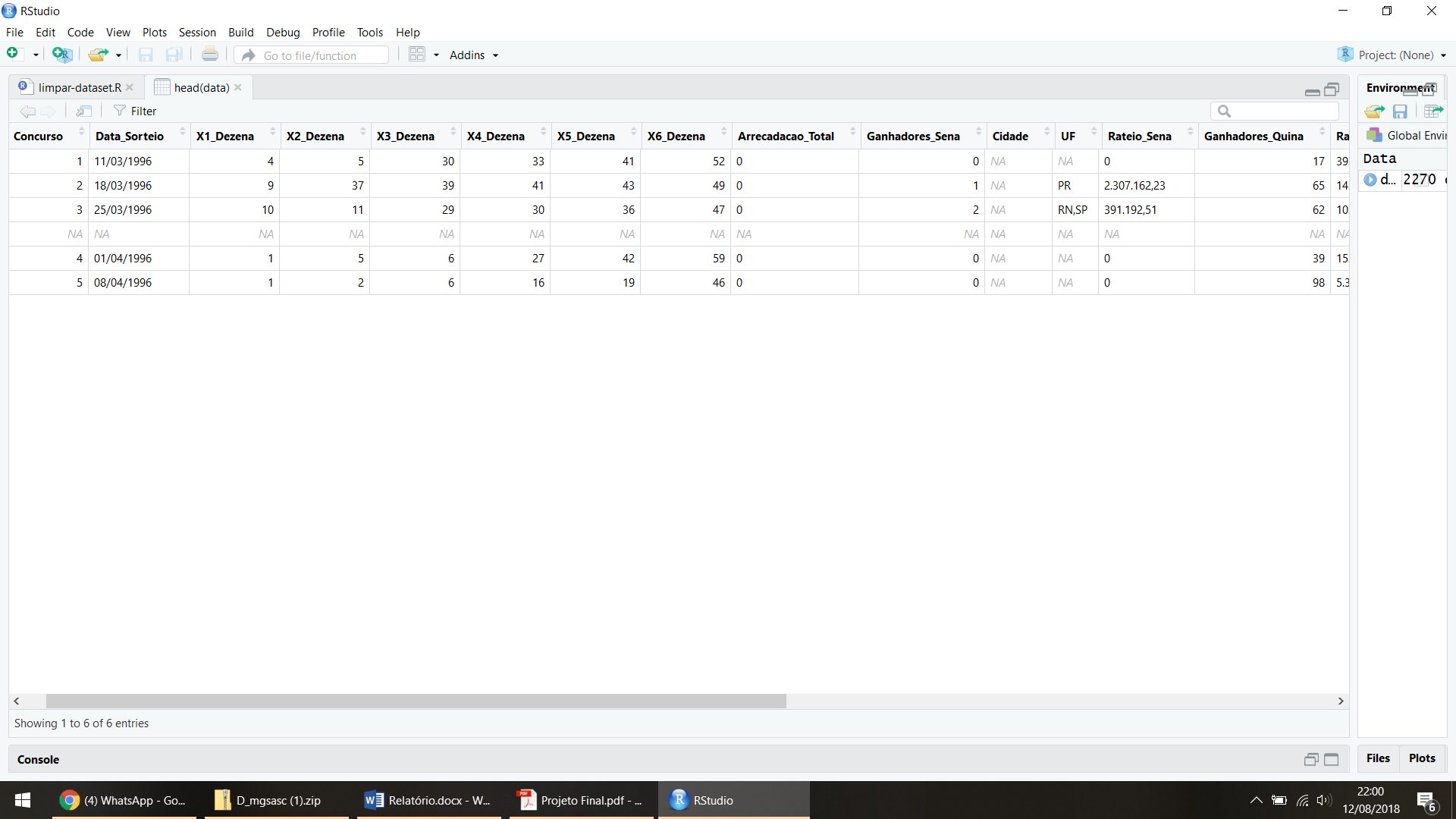
­­

Figura 2: Tabela CSV convertida.

Por fim, a remoção, substituição e formatação dos valores foram realizados para que se o manuseio com os dados fique mais coerente. A lista abaixo apresenta a ordem das operações finais da preparação dos dados.

* 1. Remoção das linhas com ***NA***;
  2. Conversão das datas formatadas como *String* para variáveis numéricas separadas (Dia, Mês e Ano);
  3. Conversão da moeda formatada como *String* para valor numérico decimal;
  4. Conversão dos valores SIM e NAO (*String*) para TRUE e FALSE (*Logical*);
  5. Agrupamento das variáveis X***n***\_Dezenas (***n*** é a ordem a qual a dezena foi sorteada) para um *dataframe* organizado com número do concurso, ordem do sorteio e dezena – figura 3;
  6. Separação dos valores da variável UF para um *dataframe* organizado com número do concurso e UF – figura 4;
  7. Criação de um *dataframe* organizado para as regiões do Brasil – figura 5;
  8. Persistência dos *dataframes* em quatro arquivos CSV.

|  |  |  |
| --- | --- | --- |
|  |  |  |
| Figura 3: clean-data-dezenas.csv | Figura 4: clean-data-estados.csv | Figura 5: clean-data-regioes.csv |

# **Análise exploratória dos dados**

Com base no *dataframe* que contém os dados gerais da Mega-Sena, podemos observar no gráfico da figura 6, o crescimento do valor do prêmio pago por ano em milhões de reais. Em 1996 o valor total que a loteria pagou como premiação foi cerca de 81 milhões de reais. Apesar da queda em 2016, o gráfico se mostra crescente chegando a pagar cerca de 1,3 bilhão de reais em 2015 – ano que mais premiou na loteria. Vale ressaltar que o valor do prêmio está relacionado ao número de apostas, sendo assim, a popularidade da Mega-Sena continua crescendo.

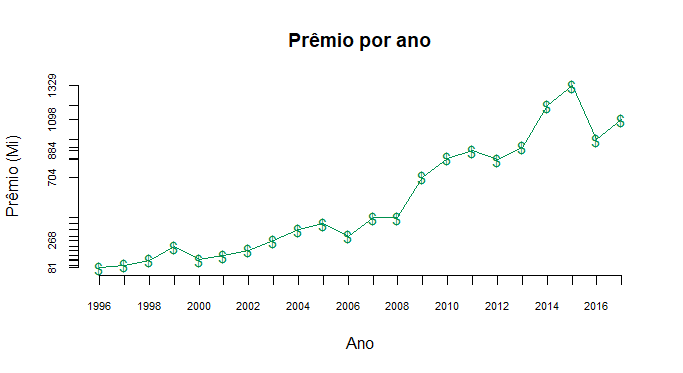


Figura 6: Prêmio por ano

Sabemos que a probabilidade de ganhar na Mega-Sena é baixíssima, porém de acordo com o contexto histórico, a figura 7 apresenta um número crescente de ganhadores. No gráfico anterior, vimos que em 2015 foi o ano com maior premiação, porém, o ano que houve mais ganhadores foi em 2012 com 54 pessoas. Isto também significa que em 2015 o valor do rateio (divisão do prêmio pelo número de ganhadores) foi um dos maiores.

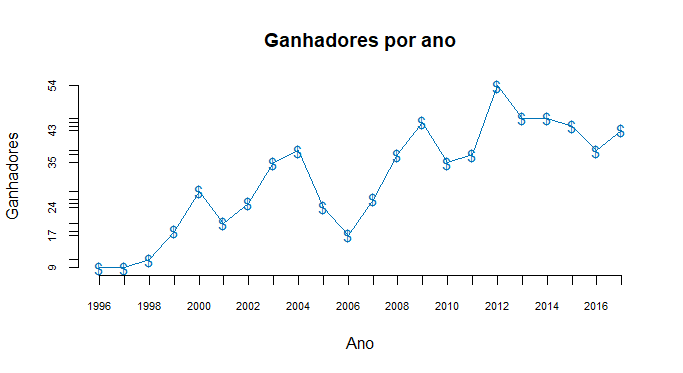


Figura 7: Ganhadores por ano

Diante de problemas atrelados à corrupção no Brasil, a confiança da população quando se envolve dinheiro em diversos setores é limitada. Na Mega-Sena existem pessoas que acreditam que os concursos são fraudados [Ref. 4]. Por ser um jogo aleatório, todas as 60 dezenas devem possuir a mesma probabilidade de serem sorteadas, resultando em uma distribuição uniforme. Para verificar parte da confiabilidade da loteria, a figura 8 apresenta o gráfico com a probabilidade da ocorrência das dezenas. Note que a linha de densidade se mantém uniforme para todas as dezenas, ou seja, a probabilidade de sorteio se mantém bem próxima para qualquer dezena. Ainda assim, existem dezenas as quais foram mais e menos sorteadas em quantidade exata. As seis mais sorteadas são: 5, 53, 10, 23, 4 e 51 respectivamente. As menos sorteadas são: 26, 55, 22, 21, 9, 25 respectivamente.

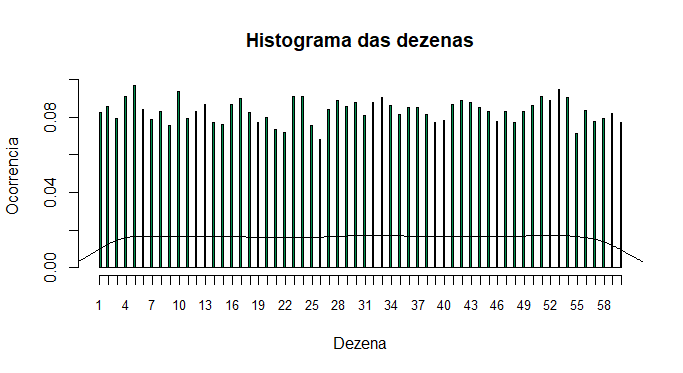


Figura 8: Histograma das dezenas

Alguns entusiastas e fanáticos pela loteria costumam elaborar fórmulas para tentar calcular o jogo que pode mudar a vida do apostador [Ref. 5]. A partir dos dados é possível verificar, na figura 9, a ocorrência das dezenas as quais foram sorteadas em concurso com ganhadores. Também é possível verificar quais foram as dezenas mais e menos apostadas entre os ganhadores. As seis mais apostadas são: 5, 10, 33, 3, 1 e 17 respectivamente. As menos apostadas são: 60, 31, 26, 55, 50, e 30 respectivamente.

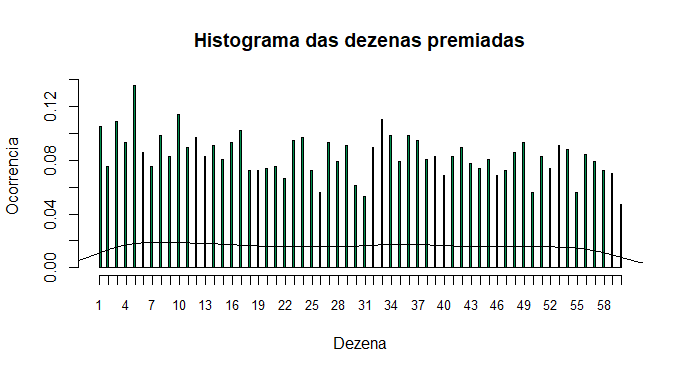


Figura 9: Histograma das dezenas premiadas

De todas as formas, seja apostando números aleatórios, apostando números combinados ou qualquer outra combinação de dezenas, apostadores da Mega-Sena possuem a esperança de ganhar. Para os pessimistas, o gráfico da figura 10 mostra que de 2000 concursos, 24% houveram ganhadores. O número de ganhadores em todos os sorteios da Mega-Sena foi 685 pessoas até o final de 2017. Também podemos observar como está o cenário dos premiados no Brasil. A figura 11 apresenta o número e a projeção de ganhadores por estados. Note que São Paulo é o estado com maior número de vencedores. Isso pode estar relacionado a cultura de apostar com frequência, como também pode estar associado ao número de lotéricas bem como a divulgação da loteria. Infelizmente os dados disponibilizados não contém a quantidade de apostas.

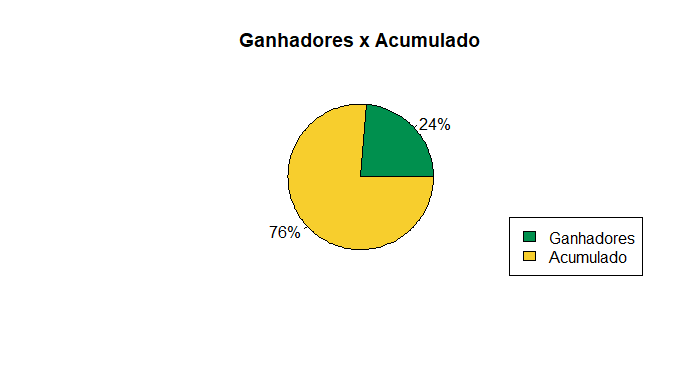


Figura 10: Ganhadores x Acumulado

|  |  |
| --- | --- |
| C:\Users\Bruno\AppData\Local\Microsoft\Windows\INetCache\Content.Word\graf-ganhadores-por-estado.png |  |
| a) Número de ganhadores | b) Projeção de ganhadores |
| Figura 11: Ganhadores por estado | |

|  |  |
| --- | --- |
| C:\Users\Bruno\AppData\Local\Microsoft\Windows\INetCache\Content.Word\graf-ganhadores-por-estado.png |  |
| a) | b) |
| Figura 11: Ganhadores por região | |

# **Conclusões**

# **Referências**

Ref. 1 – Caixa Econômica Federal – Mega-Sena.

<http://loterias.caixa.gov.br/wps/portal/loterias/landing/megasena/>

Ref. 2 – G1 – Número de ganhadores da Mega da Virada foi recorde; veja estatísticas.

<https://g1.globo.com/economia/noticia/numero-de-ganhadores-da-mega-da-virada-foi-recorde-veja-estatisticas.ghtml>

Ref. 3 – IQ Sorteios – Conheça as três loterias mais populares do Brasil.

<https://www.iq360.com.br/sorteios/artigos/as-loterias-mais-populares-do-brasil>

Ref. 4 – UOL Notícias – Coincidência inédita na Mega-Sena provoca dúvidas nas redes sociais; matemáticos explicam.

<https://noticias.uol.com.br/cotidiano/ultimas-noticias/2018/06/25/improvavel-impossivel-resultado-diferente-incentiva-teorias-de-fraude-na-mega.htm>

Ref. 5 – UOL Notícias – O que dizem os matemáticos sobre as ditas fórmulas certeiras e dicas infalíveis para se ganhar na loteria.

<https://noticias.uol.com.br/ciencia/ultimas-noticias/bbc/2017/12/29/o-que-dizem-os-matematicos-sobre-as-ditas-formulas-certeiras-e-dicas-infaliveis-para-se-ganhar-na-loteria.htm>