

EDER AUGUSTO SAKAMOTO DE SOUZA

**UMA PROPOSTA DE APLICAÇÃO DE CICLO DE VIDA DOS DADOS
PARA ANÁLISE DE DADOS METEOROLÓGICOS DO INMET**

**Monografia apresentada ao Programa de
Educação Continuada da Escola
Politécnica da Universidade de São Paulo,
para obtenção do título de Especialista,
pelo Programa de Pós-Graduação em
Engenharia de Dados e Big Data.**

EDER AUGUSTO SAKAMOTO DE SOUZA

**UMA PROPOSTA DE APLICAÇÃO DE CICLO DE VIDA DOS DADOS
PARA ANÁLISE DE DADOS METEOROLÓGICOS DO INMET**

Monografia apresentada ao Programa de Educação Continuada da Escola Politécnica da Universidade de São Paulo, para obtenção do título de Especialista, pelo Programa de Pós-Graduação em Engenharia de Dados e Big Data.

Área de concentração: Tecnologia da Informação – Engenharia/ Tecnologia/ Gestão

Orientador: Marcelo Morandini

SÃO PAULO

2024

Souza, Eder UMA PROPOSTA DE APLICAÇÃO DE CICLO DE VIDA DOS DADOS PARA ANÁLISE DE DADOS METEOROLÓGICOS DO INMET / E. Souza -- São Paulo, 2024. 42 p. Monografia (Especialização em Especialização em Engenharia de Dados e Big Data) - Escola Politécnica da Universidade de São Paulo. PECE – Programa de Educação Continuada em Engenharia. 1.Análise de Dados I.Universidade de São Paulo. Escola Politécnica. PECE – Programa de Educação Continuada em Engenharia II.t.

CURSO ENGENHARIA DE BIG DATA

Coord.: Prof. Solange N. Alves de Souza

Vice-Coord.: Prof. Anarosa Alves Franco Brandão

RESUMO

As mudanças climáticas já são visíveis e sentidas na vida das pessoas. No estado de São Paulo, especificamente, esse fenômeno se manifesta no período de chuvas, que ocorre entre outubro e março. Eventos como as inundações de São Luiz do Paraitinga (2010), a falta d'água em São Paulo (2014) e os temporais no litoral norte (2023) estão relacionados neste contexto.

Uma vez que os efeitos das mudanças climáticas se intensificarão nos próximos anos, o monitoramento das informações se torna imprescindível para o conhecimento desses fenômenos, bem como o subsídio às pessoas e órgãos competentes para o enfrentamento. Para este fim, a interpretação das bases de dados meteorológicas é fundamental para a compreensão dos fatores climáticos. Entretanto, essas informações não costumam seguir um padrão de visualização que seja comparável a outros lugares, além de apresentarem problemas de preenchimento e consistência em diferentes períodos.

Dessa forma, este trabalho realizou uma proposta de análise desses dados a partir das estações meteorológicas do estado de São Paulo (automáticas), fornecidas pelo Instituto Nacional de Meteorologia (INMET) dos últimos quinze anos. Para tanto, a análise foi apoiada na gestão do ciclo de vida dos dados da iniciativa DataONE, em que foi possível aplicar a metodologia para análise dos dados extraídos do INMET, que é a principal contribuição deste trabalho. A organização dos dados por este método de ciclo de vida proporcionou um tratamento organizado e tempestivo das informações, necessidade que se faz cada vez mais presente em análise de dados.

Palavras-chave: bases meteorológicas, qualidade de dados, ciclo de vida dos dados.

ABSTRACT

Climate change is already visible and felt in people's lives. In the state of São Paulo, specifically, this phenomenon manifests itself during the rainy season, which occurs between October and March. Events such as the floods in São Luiz do Paraitinga (2010), the lack of water in São Paulo (2014) and the storms on the north coast (2023) are related in this context.

Since the effects of climate change will intensify in the coming years, monitoring information becomes essential for understanding these phenomena, as well as providing support to the people and bodies competent to combat them. To this end, the interpretation of meteorological databases is fundamental for understanding climatic factors. However, this information does not usually follow a viewing pattern that is comparable to other places, in addition to presenting problems with completion and consistency in different periods.

Thus, this work made a proposal to analyze these data from the meteorological stations in the state of São Paulo (automatic), provided by the National Institute of Meteorology (INMET) over the last fifteen years. To this end, the analysis was supported by the management of the data life cycle of the DataONE initiative, in which it was possible to apply the methodology for analyzing data extracted from INMET, which is the main contribution of this work. The organization of data using this life cycle method provided organized and timely processing of information, a need that is increasingly present in data analysis.

Keywords: meteorological bases, data quality, data life cycle.

LISTA DE FIGURAS

| | |
|--|----|
| Figura 1: Amostra do conjunto de dados da estação automática do Mirante de Santana (INMET) | 17 |
| Figura 2: Média histórica de pluviosidade para o mês de julho em São Paulo / SP | 25 |
| Figura 3: Climograma da cidade de São Paulo (Mirante de Santana) entre 2008 e 2022 | 26 |
| Figura 4: Climograma da cidade de São Paulo (Mirante de Santana) em 2013..... | 26 |
| Figura 5: Climograma da cidade de São Paulo (Mirante de Santana) em 2014 | 27 |
| Figura 6: Climograma da cidade de São Paulo (Mirante de Santana) em 2022 | 27 |
| Figura 7: Climograma da cidade de São Luiz do Paraitinga – 2008 a 2018..... | 28 |
| Figura 8: Climograma da cidade de São Luiz do Paraitinga em 2009..... | 28 |
| Figura 9: Climograma da cidade de São Luiz do Paraitinga em 2010..... | 29 |

LISTA DE TABELAS

| | |
|---|----|
| Tabela 1: Completude dos atributos de precipitação e temperatura (máxima, média e mínima) no Mirante de Santana (2006 – 2022) | 18 |
| Tabela 2: Completude do atributo de precipitação por estação e por ano | 19 |
| Tabela 3: Tabela 3: documentação do projeto | 23 |
| Tabela 4: Documentação da tabela <i>raw</i> | 23 |

Sumário

| | | |
|-----|---|----|
| 1 | Introdução | 8 |
| 1.1 | Motivação..... | 9 |
| 1.2 | Objetivo | 9 |
| 1.3 | Justificativa | 10 |
| 1.4 | Metodologia | 10 |
| 2 | Do Ciclo de Vida à visualização | 12 |
| 2.1 | Ciclo de Vida dos Dados..... | 12 |
| 2.2 | Análise, Visualização e <i>gaps</i> de dados meteorológicos..... | 13 |
| 3 | INMET | 17 |
| 3.1 | Conjunto de dados INMET | 18 |
| 4 | Estrutura para análise dos dados | 21 |
| 4.1 | Gestão do ciclo de vida dos dados | 21 |
| 4.2 | Documentação do conjunto de dados INMET | 22 |
| 4.3 | Análise do conjunto de dados de São Paulo / SP | 24 |
| 4.4 | Análise do conjunto de dados de São Luiz do Paraitinga / SP | 27 |
| 5 | Conclusão..... | 30 |
| 5.1 | Contribuições do trabalho | 30 |
| 5.2 | Trabalhos futuros | 31 |
| | Referências Bibliográficas | 32 |
| | Apêndice: Climogramas de São Paulo / SP | 34 |
| | Apêndice: Climogramas de São Luiz do Paraitinga / SP..... | 40 |

1 Introdução

Nos últimos anos, especialmente ao longo do século XXI até a presente data, o mundo vivencia uma explosão na geração de dados dos mais diversos tipos e nos mais variados formatos. Este fenômeno exigiu mudanças, desde organizações globais até o indivíduo, no que se refere à análise exploratória dos dados em busca de boas respostas.

Nesta mudança paradigmática, em que os dados se tornam protagonistas, há o reconhecimento de compreensão, interpretação e análise de dados para que as decisões se tornem mais eficazes. Nesse sentido, são exigidas do profissional que atua na ciência da informação novas competências, que vão da coleta à habilidade em transformação em conhecimento.

A gestão do ciclo de vida dos dados reúne algumas dessas competências. Neste trabalho, a iniciativa DataONE foi utilizada para analisar dados meteorológicos obtidos de estações automáticas e cujas informações estão hospedadas no Instituto Nacional de Meteorologia (INMET). Esses dados estão relacionados a dois eventos relevantes de ordem climatológica: a seca de 2014 no município de São Paulo e a inundação de São Luiz do Paraitinga em 2010.

É prudente adiantar que análises de dados climatológicos são comuns e que as respostas são divulgadas em portais de grande mídia, inclusive. Entretanto, este trabalho consiste em utilizar uma proposta de análise apoiada em um modelo específico de gestão de ciclo de vida, uma vez que, ao permitir maior organização no processo como um todo, garante continuidade no trabalho realizado, eliminando retrabalhos.

Pretende-se, com isso, facilitar a análise dos dados climatológicos empregados no estudo e viabilizar mecanismos para que novas análises sejam realizadas com outros tipos de dados de natureza ambiental.

1.1 Motivação

Os recentes eventos em território brasileiro classificados como catástrofes climáticas merecem um olhar mais categórico e apoiado nas bases de informações disponíveis. Embora não faltem análises meteorológicas, sobretudo em relação à pluviometria e temperatura, uma estrutura mais facilitada para produzir as etapas de coleta, tratamento, documentação e divulgação de informações dos dados relacionados ao meio ambiente brasileiro pode ser aproveitada no processo.

Inicialmente, o trabalho de Morandini (2017) foi a inspiração para o aprofundamento no tema, uma vez que o autor propõe uma série de padrões de usabilidade com aplicação em temas de Mudanças Climáticas. Entre essas aplicações, há gráficos, tabelas e mapas específicos.

À medida que as pesquisas tiveram evolução para enriquecer o trabalho, a proposta do ciclo de vida dos dados elaborada por pesquisadores associados ao Projeto DataONE, conforme trabalho realizado por Anjos e Dias (2019), se mostrou satisfatória para a confecção das análises e que serão apresentadas no decorrer deste documento. Este artigo foi a base deste trabalho, cujas aplicações serão mostradas adiante.

1.2 Objetivo

O principal objetivo é promover uma análise de informações das estações automáticas, fornecidas pelo Instituto Nacional de Meteorologia (INMET), do estado de São Paulo nos últimos quinze anos. O atributo priorizado foi o da pluviometria. Com isso, a confecção de climogramas¹ a partir de informações reais e atualizadas foi elaborada para subsidiar as discussões propostas.

Para este fim, a iniciativa de apresentação de um modelo de ciclo de vida de dados foi realizada no contexto do Projeto DataOne. Embora as oito etapas propostas pela

¹ Neste trabalho, a construção da maioria dos climogramas é de curto prazo, representando o tempo de um ano específico.

iniciativa não tenham sido inteiramente apropriadas, não houve prejuízo no propósito. Espera-se que a aproximação com o método promova maior eficácia nas análises do conjunto de dados em questão, além de suportar trabalhos futuros que estão relacionados com o tema.

1.3 Justificativa

O tema das mudanças climáticas representa uma das maiores ameaças que o mundo enfrenta atualmente, com efeitos significativos para o meio ambiente, economia e vida das pessoas. Ao se apropriar dos dados relacionados ao assunto, novos modelos podem ser criados para o enfrentamento desse fenômeno, além de tempo de resposta mais curto para novas intempéries.

1.4 Metodologia

A metodologia empregada consistiu na coleta das informações do INMET, via *website* do instituto e armazenamento em banco de dados.

Como já é de amplo conhecimento, o aumento substancial de dados que há atualmente - seja por dispositivos pessoais, como aparelhos celulares, internet das coisas ou quaisquer outros equipamentos que possuam conexão com a internet - diz respeito ao fenômeno *Big Data* que possui três características fundamentais (os três “Vs”): volume, velocidade e variedade. Conforme apontam Anjos e Dias, “Volume está relacionado à quantidade avassaladora de dados que são gerados diariamente. Variedade é correlacionada à diversidade, heterogeneidade desses dados. E velocidade em relação à rapidez em que esses dados são disponibilizados...” (2019, p.81).

Nesse sentido, diferentes possibilidades são criadas e novas formas de trabalhar com estes dados são exigidas. A curadoria de dados passa a ser essencial para o êxito das análises de dados no contexto do Big Data. O processo escolhido para este trabalho foi o Ciclo de Vida dos Dados – DataOne, bem como as suas etapas: planejar, coletar, assegurar, descrever, preservar, descobrir, integrar, analisar.

Buscando validar computacionalmente as análises, foi utilizado um SGBD (SQL Server, versão 18.12.1) para o armazenamento, processamento e criação de consultas que forneçam a base para analisar as informações. A escolha do SGBD se deu em virtude da suficiência da ferramenta para armazenamento e tratamento desta proposta.

2 Do Ciclo de Vida à visualização

2.1 Ciclo de Vida dos Dados

O ciclo de Vida dos Dados - Dataone tem como missão “permitir a criação de novas ciências e conhecimento através do acesso universal aos dados sobre a vida na terra e o ambiente que a sustenta” (DataONE apud. Anjos e Dias, 2019, p.84). Essa iniciativa permite uma visão das etapas que visam a preservação de dados e seu gerenciamento. “Ressalta-se que existem variações do ciclo de vida dos dados de acordo com práticas, domínios e necessidades de cada comunidade” (p.,84).

Ainda segundo Anjos e Dias (2019), o Ciclo de Vida dos Dados – DataONE é composto por oito etapas, a saber:

- 1) **Planejar**, que corresponde a um planejamento de como irá ocorrer o processo. Aqui, todos os procedimentos metodológicos e etapas são pensados. Um exemplo é como se dará a coleta dos dados, onde serão armazenados, como serão organizados, como serão gerenciados, como serão descritos, como serão compartilhados, como serão preservados e qual o esforço orçamentário envolvido. (p.85)
- 2) **Coletar**, que consiste onde e como os dados serão coletados. A recomendação dos autores é criar um modelo para coleta, descrever o conteúdo, organizar em arquivo (que podem ser planilhas ou banco de dados), utilizar o mesmo formato de todos em todo o arquivo (e, assim, garantir padrão), uso de caracteres de texto simples, utilização de softwares e hardwares livres, atribuição de nomes de arquivos descritivos, manutenção dos dados crus (*raw layer*), criação de tabelas de parâmetros. (p.86)
- 3) **Assegurar**, que é a adoção de procedimentos e critérios que asseguram a qualidade e validação dos dados. São tarefas essenciais nesta etapa: descrição das condições na etapa de coleta, verificação a partir de conjuntos similares, verificação de dados manualmente inseridos, elaboração de resumos estatísticos e gráficos, que evidenciarão valores improváveis, faltantes ou impossíveis. (p.86)

- 4) **Descrever**, que contempla uma descrição detalhada por meio de metadados, que diz respeito à documentação completa dos dados, além das seguintes descrições: contexto digital, pessoal e partes interessadas, contexto científico. Adicionalmente, a etapa considera o fornecimento de informações sobre os parâmetros e informar como deve ser a citação do conjunto de dados. (p.86)
- 5) **Preservar**, que é a etapa de planejamento de preservação de curto e longo prazos. É essencial que o trabalho possua um repositório de dados que forneça orientações, como gerar metadados, como preservar e quais formatos utilizar, por exemplo. Outros elementos são o uso de terminologias padrões, políticas gerais, atribuição, pessoal responsável pelo conjunto de dados, contexto do conjunto de dados, histórico de revisão, links para dados de origem, como citar o conjunto de dados e direitos de propriedade intelectual. (p.87).
- 6) **Descobrir**, que é a identificação de outros conjuntos que podem agregar valor a um determinado trabalho ou pesquisa, além de fazer com que os dados atuais, já explorados, possam contribuir com outros trabalhos. Para tal, faz-se necessária a cultura de compartilhamento e também a publicidade do trabalho realizado. (p.87)
- 7) **Integrar**, que é a integração entre o conjunto de dados da pesquisa a outros recursos. Para isso, os dados devem ser compatíveis e comparáveis (p.87).
- 8) **Analisar**, que é a exploração dos dados, análise e visualização. Para que esta etapa seja possível, é necessária a documentação dos métodos e etapas (p.88).

2.2 Análise, Visualização e *gaps* de dados meteorológicos

O ponto de partida da visualização de dados é a análise de dados sob a perspectiva da história. Silva (2019) promove uma leitura sobre visualização de dados desde a sua origem. Além disso, o autor traz importantes definições sobre o tema, onde “o principal objetivo da visualização de dados é comunicar informações ou ideias complexas de forma clara, precisa e eficiente, de uma forma que ajude os usuários a analisar e raciocinar sobre dados e evidências” (p.207). Desta maneira, este objetivo está diretamente relacionado com a proposta deste estudo, onde se pretende a facilitação na leitura dos dados meteorológicos através de uma visualização intuitiva.

Ainda que Silva faça uma análise histórica - desde a pré-história - o foco deste trabalho está nos anos 1990 em diante.

Ao longo de seu trabalho, Silva coloca a visualização de dados no contexto do Big Data. Sobre este aspecto, o autor diz:

A velocidade com que os dados são gerados aumentou, impulsionada por uma economia crescente baseada em informações. Os dados criados pela atividade da Internet e um número crescente de sensores no ambiente, como satélites e câmeras de tráfego, são conhecidos como "Big Data". Processamento, análise e comunicação desses dados apresentam uma variedade de desafios éticos e analíticos para a visualização de dados (MENDONÇA apud. Silva, 2010, p.207).

Em resumo, o *Big Data* não consegue contribuir de forma relevante sem um olhar criterioso para a visualização. Desta maneira, o autor estabelece um guia de quatro passos, que este trabalho tentará seguir (p.218):

- Ter um conjunto de dados limpo: significa um conjunto de dados no formato apropriado para a maioria das ferramentas de visualização
- Única mensagem de comunicação: selecionar uma única mensagem, que será destacada em cada gráfico. Dependendo do público, o tipo de gráfico e o nível de precisão necessário.
- Escolher o gráfico adequado: A forma de um gráfico depende do que queremos transmitir: por exemplo, comparar, distribuir etc.
- Design e cor: destacar o que é essencial, usando cores.

Sobre os objetivos de uma visualização iterativa, Silva elenca três (p.219): Colocar os dados em contexto; que o usuário seja capaz de explorar os dados; que o usuário tenha a possibilidade de encontrar padrões.

Morandini (2017), embora foque sua obra em aplicações de software, dialoga com a necessidade e benefícios da visualização de dados, uma vez que “Gráficos são recursos muito visuais muito aplicados em Sistemas de Visualização de Informação devido à sua capacidade de fornecer a informação de forma visual com cores, formas e desenhos” (p.14). O autor destaca, porém, que os gráficos carecem de informações sobre como se deu a criação da informação, bem como quais dados foram utilizados

para tal análise. Nesse sentido, Morandini aponta para importância das tabelas, dado que:

A visualização de informação por si só pode não ser suficiente para proporcionar ao usuário a totalidade da compreensão dos fatos. Um conjunto de dados, responsável por apresentar informações, pode sofrer com alterações temporais, comportamentais e conceituais. Desta forma, conhecer a informação e os fatores que permitiram suas construções deve ser uma constante para desenvolvedores. Assim, permitir que desenvolvedores de IHC utilizem tabelas em seus sistemas e que essas tabelas possam efetivamente auxiliar o usuário na visualização, análise e compreensão da informação. Destaca-se como fundamental a apresentação das fontes provenientes dos dados tabulados. (MORANDINI, 2017, p.12.).

Ainda que as tabelas não forneçam um atrativo visual, segundo o autor, elas permitem uma maior riqueza de detalhes, sendo ideal nos casos de necessidade de análises mais detalhadas de fenômenos.

Outra frente relevante diz respeito às falhas de preenchimento. A literatura especializada é praticamente unânime em apontar a incompletude dos dados como um dos principais problemas em análise de dados meteorológicos, como apontam Bier e Ferraz (2015): “Um dos grandes problemas que surgem ao se trabalhar com dados medidos em estações meteorológicas é a incompletude destes”. É importante salientar que esta é uma questão que não surgiu nos últimos anos. Já em 2005, Jurandir Zullo Junior e colegas já anunciavam para a necessidade de consistência dos dados meteorológicos: “A disponibilidade de séries históricas de dados pluviométricos diários consistidos, com qualidade e confiabilidade, é fundamental para a realização de trabalhos na área de modelagem agroclimatológica”. Os autores realizaram uma análise sobre as informações de precipitação desde 1934 e divulgaram critérios de inconsistência das informações, conforme segue: “pelo menos dois meses sem dados; pelo menos dois meses da estação chuvosa (outubro a março) sem registro de chuvas; c) pelo menos um mês sem dados e um mês da estação chuvosa sem registro de chuvas”.

Sabino e Souza (2022), ao realizarem um trabalho de apresentação de software capaz de preenchimento dos dados faltantes nas estações automáticas do estado de Mato Grosso, entre 2008 e 2020, chegaram a um resultado preliminar em que, dentre as 33

estações meteorológicas pesquisadas, os *gaps* ocorrem em 18,61%, em média, para cada estação e o máximo de dados faltantes chega a 62% (Radiação Solar Descendente na estação de São Félix do Araguaia / MT).

Conforme aponta Carpenedo et.al (2022), assim como a previsão do tempo, o monitoramento é uma ferramenta fundamental para tomadas de decisão derivadas do tema. E mesmo assim, “o país apresenta carência de séries temporais de dados meteorológicos longos e homogêneos” (p.28). No trabalho elaborado pelos autores, focado no possível preenchimento de dados - a partir de outras cinco bases disponíveis - não preenchidos de pluviometria em uma estação automática de Uberlândia / MG, embora a operação estivesse operando desde o final de 2002, houve falta de informações em quase 44% dos dados até o fim de 2008, o que impediu a análise ao longo desse período. Os resultados da pesquisa apontam que não foi possível chegar a resultados satisfatórios ao utilizar informações de outras bases. Portanto, a preocupação com a qualidade dos dados pluviométricos é um tema transversal a qualquer trabalho que envolva essas informações.

3 INMET

O Instituto Nacional de Meteorologia (INMET) é uma instituição brasileira de monitoramento e pesquisa meteorológica. Fundado em 1909, o INMET tem desempenhado um papel importante na coleta de dados, previsão do tempo e estudos climáticos, com contribuições para a compreensão dos fenômenos atmosféricos no Brasil.

Uma das principais responsabilidades do INMET é fornecer informações atualizadas sobre as condições meteorológicas em todo o território nacional. Através de sua rede de estações meteorológicas, o instituto coleta dados sobre temperatura (máxima, média e mínima), pluviosidade, umidade, pressão atmosférica, ventos, entre outros parâmetros em estações automáticas e convencionais. Essas informações são cruciais para a previsão do tempo, análise dos chamados “desastres naturais” (principalmente aqueles relacionados a secas e inundações) e a tomada de decisões em determinados setores da economia (por exemplo, a agricultura e a aviação).

Além disso, o INMET possui importância na pesquisa científica, promovendo estudos avançados sobre o clima brasileiro. Suas pesquisas são fundamentais para compreender as mudanças climáticas, identificar tendências de longo prazo e desenvolver estratégias de adaptação. Esse papel é objeto de estudo deste trabalho, no sentido de procurar maximizar essa função.

O instituto também desempenha uma função na educação, pois compartilha informações meteorológicas de maneira acessível para o público em geral. Ou seja, as informações são abertas e relativamente fáceis de serem consultadas. Por conta disso, a ideia aqui é de, ao utilizar os dados dos bancos de dados do INMET, tornar a informação, não apenas acessível, mas também compreensível em vista da contribuição de uma sociedade mais preparada diante das condições meteorológicas adversas.

3.1 Conjunto de dados INMET

O INMET permite realizar o download dos dados das estações automáticas ou convencionais e de todas as regiões do Brasil. Os dados utilizados neste trabalho possuem as seguintes características:

- Região Sudeste (estado de São Paulo, apenas)
- Estações automáticas
- Dados diários

Dentro dessas características, as estações meteorológicas do INMET contêm os seguintes atributos:

- Precipitação Total
- Pressão Atmosférica
- Temperatura do ponto de orvalho média
- Temperatura máxima
- Temperatura média
- Temperatura mínima
- Umidade Relativa do ar média
- Umidade Relativa do ar mínima
- Vento, rajada máxima
- Vento, rajada média

Dentro deste conjunto de atributos, a **Precipitação Total** foi priorizada, uma vez que é o fenômeno de maior impacto na sociedade. Além da precipitação, serão trabalhadas as temperaturas (máxima, média e mínima) por serem indicadores de referência em assuntos climáticos.

Abaixo, segue um exemplo de estrutura extraída do INMET:

Figura 1: Amostra do conjunto de dados da estação automática do Mirante de Santana (Fonte: INMET)

| Nome: SAO PAULO - MIRANTE | | | | | | | | | | | | | |
|----------------------------------|----------------------------|---------------------------|---------------------------|----------------------------|---------------------------|----------------------------|----------------------------|----------------------------|-----------------------------|--------------------------------------|------|--|--|
| Codigo Estacao: A701 | | | | | | | | | | | | | |
| Latitude: -23.19638888 | | | | | | | | | | | | | |
| Longitude: -46.61999999 | | | | | | | | | | | | | |
| Altitude: 785.64 | | | | | | | | | | | | | |
| Situacao: Operante | | | | | | | | | | | | | |
| Data Inicial: 2006-07-24 | | | | | | | | | | | | | |
| Data Final: 2023-04-26 | | | | | | | | | | | | | |
| Periodicidade da Medicao: Diaria | | | | | | | | | | | | | |
| Data Medicao | PRECIPITACAO TOTAL, DIARIO | PRESSAO ATMOSFERICA MEDIA | TEMPERATURA DO PONTO DE C | TEMPERATURA MAXIMA, DIARIA | TEMPERATURA MEDIA, DIARIA | TEMPERATURA MINIMA, DIARIA | UMIDADE RELATIVA DO AR, ME | UMIDADE RELATIVA DO AR, MI | VENTO, RAJADA MAXIMA DIARIA | VENTO, VELOCIDADE MEDIA DIARIA (AUT) | | | |
| 24/07/2006 | null | null | null | null | null | null | null | null | null | null | | | |
| 25/07/2006 | null | null | null | null | null | null | null | null | 45 | 7,2 | null | | |
| 26/07/2006 | 0 | 925,2875 | 13,55 | 27,5 | 21,1 | 15,8 | 64 | 44 | 10,1 | 1,625 | | | |
| 27/07/2006 | 0,4 | 925,191667 | 13,391667 | 28,2 | 21,454167 | 14,4 | 61,833333 | 42 | 8,4 | 1,566667 | | | |
| 28/07/2006 | 0 | 924,483333 | 16,366667 | 29,4 | 22,654167 | 17,5 | 69,541667 | 51 | 9,1 | 1,6875 | | | |
| 29/07/2006 | 8,2 | 924,058333 | null | 24,8 | 18,1625 | 17,7 | null | 60 | 11,5 | 2,820833 | | | |
| 30/07/2006 | 0,8 | 929,495833 | null | 13,3 | 11,795833 | 11,4 | null | null | 7,6 | 2,5625 | | | |
| 31/07/2006 | 13,8 | 932,195833 | null | 15,9 | 11,254167 | 10,5 | null | 6,3 | 2,070833 | | | | |
| 01/08/2006 | 2,6 | 932,8375 | null | 15,2 | 13,345833 | 11,9 | 97,5 | 94 | 5,5 | 1,245833 | | | |
| 02/08/2006 | 0,2 | 932,766667 | null | 15,4 | 14,070833 | 13,2 | 98,866667 | 94 | 5,4 | 1,866667 | | | |
| 03/08/2006 | 0 | 931,3625 | 14,433333 | 19,8 | 15,445833 | 13,4 | 94,166667 | 76 | 5,9 | 2 | | | |
| 04/08/2006 | 0 | 929,843833 | 14,702625 | 24,1 | 17,829167 | 13,9 | 80 | 51 | 4,9 | 3,1875 | | | |
| 05/08/2006 | 0 | 929,641667 | 15,772222 | 26,2 | 18,4625 | 12,9 | 78,5 | 50 | 4,6 | 1,0875 | | | |
| 06/08/2006 | 0 | 929,0125 | 14,670833 | 28,6 | 20,804167 | 13,2 | 71,5 | 41 | 6,9 | 1,075 | | | |
| 07/08/2006 | 0 | 928,679167 | 13,741667 | 28,5 | 21,508333 | 15,1 | 64,083333 | 35 | 6,7 | 1,370833 | | | |
| 08/08/2006 | 0 | 928 | 13,545833 | 28,6 | 21,75 | 15,5 | 61,75 | 38 | 6,9 | 1,633333 | | | |
| 09/08/2006 | 0 | 926,45 | 10,875 | 27,3 | 22,091667 | 15,7 | 49,708333 | 35 | 11,2 | 2,2125 | | | |
| 10/08/2006 | 0 | 927,1625 | 13,2875 | 29,2 | 23,425 | 19,7 | 53,416667 | 41 | 8,3 | 1,895833 | | | |
| 11/08/2006 | 0 | 928,5 | 15,258333 | 28,2 | 21,570833 | 16,9 | 69,875 | 42 | 7,9 | 2,1375 | | | |
| 12/08/2006 | 0 | 927,633333 | 14,5875 | 28,4 | 22,208333 | 16,3 | 64,4 | 41 | 8,5 | 1,570833 | | | |
| 13/08/2006 | 0 | 927,245833 | 14,433333 | 28,4 | 22,104167 | 16,7 | 63,666667 | 44 | 8,6 | 1,429167 | | | |
| 14/08/2006 | 0 | 925,433333 | 14,033333 | 28,4 | 22,583333 | 16,6 | 60,125 | 42 | 11,1 | 1,9 | | | |
| 15/08/2006 | 0 | 925,490952 | 14,492381 | 30,1 | null | 18 | 59,857143 | 39 | 6,8 | 1,171429 | | | |
| 16/08/2006 | null | 925,733333 | 16,55 | 30,7 | 22,758333 | 17,5 | 72,458333 | 38 | 10,3 | 2,35 | | | |
| 17/08/2006 | 0 | 928,0625 | 17,566667 | 23,3 | 19,191667 | 17,8 | 90,583333 | 73 | 8,6 | 2,325 | | | |
| 18/08/2006 | 0 | 927,8375 | 17,4 | 24,5 | 18,729167 | 16,7 | 92,458333 | 69 | 7,1 | 2,479167 | | | |
| 19/08/2006 | 0 | 925,866667 | 16,625 | 19 | 17,6125 | 17,1 | 94,125 | 86 | 8,1 | 2,0125 | | | |
| 20/08/2006 | 0 | 924,7375 | 14,75 | 22,5 | 17,570833 | 15,2 | 85,333333 | 51 | 9 | 1,808333 | | | |
| 21/08/2006 | 0,6 | 928,916667 | 10,1625 | 17,9 | 15,1125 | 11 | 82,916667 | 55 | 8,6 | 2,1875 | | | |

Não configura exatamente uma novidade concluir que os dados dispostos dessa forma não trazem respostas muito efetivas. Dessa maneira, a seguir serão apresentadas algumas análises deste conjunto de informações (Estação automática do Mirante de Santana – 24/07/2006 a 26/04/2023), com ênfase para os atributos de interesse deste trabalho.

Há, no estado de São Paulo, 44 estações automáticas com dados disponíveis. Entretanto, 5 delas não estão ativas. Portanto, o total da análise é de um total de 39 estações meteorológicas, que são as ativas em abril/2023.

A tabela abaixo corresponde ao montante de 6.121 registros da estação automática do Mirante de Santana. Cada registro corresponde a um dia de dados (24/07/2006 a 26/04/2023). Aqui pode-se notar a relativa alta completude dos atributos, acima de 99% em todos eles. Entretanto, isso não ocorre nas demais estações, como pode ser verificado na tabela 2 – agregada por ano - onde, no atributo de precipitação, há perdas visíveis de dados em determinados períodos. Pode-se destacar o provável

efeito que a pandemia de COVID-19 teve na captura das informações, dado a falta concentrada em diversas estações entre os anos de 2020 e 2022.

Tabela 1: Completude dos atributos de precipitação e temperatura (máxima, média e mínima) no Mirante de Santana (2006 – 2022). Elaborada pelo autor.

| São Paulo - Mirante | registros | precip | tmax | tmed | tmin |
|---------------------|-----------|--------|-------|-------|-------|
| Preenchidos | 6.121 | 6.082 | 6.092 | 6.072 | 6.096 |
| faltantes | - | 39 | 29 | 49 | 25 |
| % preenchimento | - | 99,4% | 99,5% | 99,2% | 99,6% |

Tabela 2: Completude do atributo de precipitação por estação e por ano. Elaborada pelo autor

| Município | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 | 2021 | 2022 | Média de Completude (2008-22) |
|------------------------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|-------------------------------|
| São Paulo - Mirante | 99% | 99% | 100% | 99% | 100% | 99% | 97% | 100% | 99% | 99% | 100% | 100% | 100% | 100% | 100% | 99% |
| Presidente Prudente | 100% | 99% | 93% | 100% | 100% | 100% | 100% | 99% | 80% | 100% | 100% | 100% | 100% | 99% | 75% | 96% |
| São Paulo - Interlagos | | | | | | | | | | | 80% | 100% | 100% | 100% | 95% | 95% |
| São Carlos | 100% | 99% | 100% | 100% | 81% | 100% | 75% | 100% | 88% | 100% | 84% | 100% | 100% | 99% | 82% | 94% |
| Pradópolis | 69% | 99% | 100% | 87% | 100% | 99% | 83% | 100% | 83% | 100% | 80% | 100% | 75% | 91% | 100% | 91% |
| Marília | | | | | | | | | | 62% | 100% | 81% | 100% | 100% | 100% | 90% |
| Bauru | 100% | 43% | 86% | 99% | 100% | 79% | 88% | 100% | 92% | 100% | 87% | 100% | 100% | 96% | 75% | 90% |
| Itapeva | 100% | 95% | 68% | 97% | 58% | 100% | 96% | 85% | 100% | 100% | 100% | 100% | 79% | 75% | 79% | 89% |
| Franca | 98% | 21% | 99% | 99% | 100% | 99% | 67% | 100% | 100% | 66% | 100% | 65% | 87% | 97% | 99% | 86% |
| São Simão | | | | | | | | | | | | 49% | 100% | 98% | 98% | 86% |
| Campos do Jordão | 67% | 48% | 100% | 100% | 100% | 100% | 100% | 100% | 95% | 80% | 100% | 100% | 84% | 71% | 44% | 86% |
| Sorocaba | 87% | 99% | 99% | 92% | 96% | 96% | 96% | 98% | 97% | 100% | 90% | 81% | 81% | 6% | 68% | 86% |
| Lins | 100% | 99% | 78% | 100% | 86% | 58% | 76% | 85% | 100% | 100% | 86% | 80% | 84% | 83% | 67% | 86% |
| Valparaíso | 100% | 42% | 76% | 91% | 100% | 100% | 81% | 100% | 100% | 92% | 87% | 100% | 77% | 79% | 58% | 85% |
| Iguape | 58% | 89% | 92% | 90% | 100% | 96% | 75% | 86% | 87% | 79% | 93% | 100% | 100% | 46% | 76% | 85% |
| Tupã | | | | | | | | | | 59% | 90% | 83% | 99% | 93% | 82% | 85% |
| Casa Branca | 100% | 99% | 100% | 83% | 66% | 98% | 88% | 90% | 92% | 78% | 82% | 90% | 48% | 55% | 98% | 85% |
| Barretos | | | 53% | 100% | 100% | 100% | 91% | 100% | 100% | 100% | 85% | 58% | 100% | 65% | 46% | 84% |
| Jose Bonifácio | 100% | 99% | 89% | 99% | 100% | 100% | 100% | 85% | 100% | 86% | 86% | 77% | 27% | 9% | 90% | 83% |
| São Luiz do Paraitinga | 100% | 99% | 100% | 100% | 87% | 98% | 90% | 98% | 100% | 100% | 100% | 33% | 58% | 0% | 74% | 82% |
| Itapira | 95% | 89% | 99% | 99% | 84% | 100% | 100% | 94% | 83% | 100% | 100% | 84% | 38% | 0% | 64% | 82% |
| São Miguel Arcanjo | 80% | 81% | 58% | 72% | 95% | 69% | 62% | 100% | 98% | 100% | 100% | 66% | 100% | 74% | 60% | 81% |
| Piracicaba | 100% | 99% | 97% | 59% | 100% | 100% | 85% | 96% | 72% | 87% | 100% | 82% | 54% | 0% | 81% | 81% |
| Barueri | | | | 76% | 97% | 100% | 100% | 100% | 88% | 100% | 75% | 80% | 50% | 9% | 87% | 80% |
| Taubaté | 60% | 55% | 63% | 86% | 87% | 83% | 100% | 83% | 100% | 81% | 100% | 100% | 100% | 16% | 87% | 80% |
| Rancharia | 100% | 98% | 98% | 54% | 80% | 40% | 67% | 99% | 95% | 66% | 93% | 98% | 88% | 50% | 68% | 80% |
| Dracena | | | | | | | | | 14% | 100% | 100% | 96% | 41% | 100% | 99% | 79% |
| Ibitinga | 100% | 76% | 78% | 99% | 100% | 88% | 99% | 100% | 100% | 63% | 77% | 76% | 8% | 21% | 84% | 78% |
| Jales | 100% | 99% | 84% | 61% | 100% | 100% | 100% | 100% | 75% | 100% | 86% | 80% | 55% | 2% | 25% | 78% |
| Ituverava | 29% | 35% | 100% | 100% | 100% | 64% | 93% | 89% | 100% | 92% | 100% | 92% | 64% | 2% | 84% | 76% |
| Bragança Paulista | | | | | | | | | | 3% | 99% | 74% | 100% | 85% | 84% | 74% |
| Barra Bonita | 56% | 99% | 100% | 100% | 100% | 100% | 100% | 95% | 100% | 83% | 73% | 8% | 0% | 0% | 68% | 72% |
| Airanhã | 98% | 75% | 75% | 95% | 94% | 100% | 94% | 72% | 100% | 79% | 75% | 62% | 7% | 0% | 24% | 70% |
| Cachoeira Paulista | | | | | | | | | | 19% | 100% | 100% | 83% | 0% | 86% | 65% |
| Avaré | 98% | 41% | 71% | 70% | 85% | 53% | 36% | 100% | 100% | 99% | 65% | 90% | 0% | 0% | 59% | 64% |
| Barra do Turvo | 39% | 8% | 90% | 58% | 25% | 74% | 82% | 89% | 100% | 100% | 91% | 80% | 0% | 0% | 75% | 61% |
| Bertioga | | | | | | | | | | 91% | 100% | 90% | 0% | 0% | 79% | 60% |
| Registro | | | | | | | | | | 86% | 99% | 91% | 0% | 0% | 74% | 58% |
| Bebedouro | | | | | | | | | 16% | 100% | 87% | 91% | 30% | 31% | 27% | 55% |

4 Estrutura para análise dos dados

4.1 Gestão do ciclo de vida dos dados

Para realizar as análises, buscou-se a aplicação das etapas descritas anteriormente a respeito da gestão do ciclo de vida dos dados, sob a perspectiva da iniciativa DataOne, conforme detalhamento a seguir:

- **Planejamento:** Conforme compete a essa etapa, a coleta dos dados possui o caminho através do próprio *website* do INMET. Foram selecionados o tipo de estação (automática ou convencional) e os indicadores. A partir disso, uma requisição via *e-mail* foi gerada e os dados foram recepcionados, também via *e-mail*, no endereço da requisição. Quanto ao armazenamento, foi utilizado um SGBD (SQL Server v18.12.1), por se mostrar mais adequado em virtude do volume das informações (ou seja, o SQL SERVER comporta a massa de dados extraída).

Em relação ao formato, todos os campos estão em formato VARCHAR (similar a STRING, um campo no formato de texto). São poucos atributos, então a modelagem física não é um risco de alta severidade.

Quanto aos demais elementos desta etapa, segue: o responsável pelos dados é o autor deste trabalho; foi elaborado um documento contendo a descrição para facilitar sua preservação; não há intencionalidade de compartilhamento dos dados (a não ser em caso de continuidade do trabalho).

- **Coleta:** A coleta dos dados ocorreu por meio de requisição no *website* do INMET, como registrado anteriormente. Esta forma de coleta é relativamente simples. A descrição dos arquivos está na documentação do conteúdo, tema que também foi adiantado na etapa de planejamento. Para o armazenamento em si, feito localmente, consta: a versão cru dos dados (o próprio *output* do INMET, sem alteração ou tratamento, em uma única pasta específica em diretório *WINDOWS*; a replicação desses dados crus em uma única tabela hospedada no SQL SERVER; criação de *views* que corresponderão à etapa intermediária de tratamento (camada *trusted*) e, finalmente, uma tabela final

que será a camada *analytics* ou *refined*, que geraram as respostas e será a tabela final que poderá, inclusive, se conectar com ferramentas de B.I (*Power B.I*, por exemplo). Neste estudo, não é necessário que camada *refined* seja importada para uma ferramenta de B.I, embora o intuito seja de tornar essa possibilidade válida.

- **Etapas de Assegurar:** Algo que é muito comum neste tipo de informação é a ausência de dados pelos mais variados fatores. Esses fatores não são objeto de estudo deste trabalho. Entretanto, a validação dos dados faltantes nos atributos de temperatura (média, mínima e máxima) e precipitação constam na tabela 2.
- **Descrição:** A documentação do conjunto de dados está em tópico específico.
- **Preservação:** Para este trabalho, não foi necessário um plano de preservação. Em caso de continuidade, um novo planejamento pode ser realizado, visando um plano de preservação de curto e longo prazos.
- **Descobrimeto:** Assim como na preservação, esta etapa ainda não foi necessária para o trabalho. Contudo, em caso de novas possibilidades, a devida publicidade será realizada, dentro dos trâmites acadêmicos estabelecidos.
- **Integração:** A escolha do SGBD para armazenamento e análise dos dados reside na ideia de que o recurso é, além de amplo conhecimento pelos profissionais da ciência da informação, integrável a outros bancos de dados. Por isso, a modelagem lógica e física possui é relativamente simples, a fim de tornar a integração possível, exigindo menos tarefas adaptativas.
- **Análise:** A exploração, análise e respostas referente ao conjunto de dados utilizado está nas partes 4.3 e 4.4 deste capítulo.

4.2 Documentação do conjunto de dados INMET

Conforme disposto no capítulo anterior, a documentação exposta adiante se propõe a fornecer a descrição do projeto e do conjunto de dados em si.

Nesta etapa, faz sentido descrever o contexto digital, pessoal e partes interessadas e o contexto científico, além do conjunto de dados. Informações sobre os parâmetros já

foram fornecidas anteriormente e, nesta versão, não é relevante o fornecimento de forma de citação do conjunto de dados trabalhados.

Este capítulo foi dividido em duas partes, sendo:

- Documentação do Projeto, onde a descrição se concentra no contexto do projeto como um todo
- Conjunto de dados, onde se tem uma documentação mais clássica do significado dos campos da tabela *raw* no SQL SERVER

4.2.1 Documentação do projeto

Tabela 3: documentação do projeto. Elaborada pelo autor

| tipo | descricao |
|---------------------|--|
| Contexto Digital | Big data, dados como protagonistas dos insights, necessidade de aprimoramento das análises em busca de boas respostas |
| Pessoal | autor do trabalho |
| Partes Interessadas | autor do trabalho, o PECE/POLI e demais pessoas interessadas no tema |
| Contexto Científico | Especialização em engenharia de dados e Big Data. Trabalho de Conclusão de Curso |
| Relevância | A análise de dados de dados meteorológicos é um tema de importância na atualidade, pois permite respostas mais tempestivas em um contexto de mudanças climáticas |

4.2.2 Tabela *raw* no SQL Server (conjunto de dados)

nome da tabela: estudoinmet.dbo.estacoes_automaticas_raw

Tabela 4: Documentação da tabela raw. Elaborada pelo autor

| campo_inmet | campo_analise | descricao | período | elegível | formato_analise |
|--|---------------|---|---------|----------|-----------------|
| Data Medicao | dtRegistro | data da medição do indicador | diário | 1 | varchar |
| PRECIPITACAO TOTAL, DIARIO (AUT)(mm) | na | precipitação total | diário | 1 | na |
| PRESSAO ATMOSFERICA MEDIA DIARIA (AUT)(mB) | na | valor da pressão atmosférica | diário | 0 | na |
| TEMPERATURA DO PONTO DE ORVALHO MEDIA DIARIA | na | temperatura média do ponto de orvalho no período | diário | 0 | na |
| TEMPERATURA MAXIMA, DIARIA (AUT)(°C) | tmax | Valor de Temperatura máxima no período | diário | 1 | varchar |
| TEMPERATURA MEDIA, DIARIA (AUT)(°C) | tmed | Valor de Temperatura média no período | diário | 1 | varchar |
| TEMPERATURA MINIMA, DIARIA (AUT)(°C) | tmin | Valor de temperatura mínima no período | diário | 1 | varchar |
| UMIDADE RELATIVA DO AR, MEDIA DIARIA (AUT)(%) | na | Umidade relativa do ar média para o período | diário | 0 | na |
| UMIDADE RELATIVA DO AR, MINIMA DIARIA (AUT)(%) | na | Umidade relativa do ar mínima para o período | diário | 0 | na |
| VENTO, RAJADA MAXIMA DIARIA (AUT)(m/s) | na | rajada máxima do vento para o período | diário | 0 | na |
| VENTO, VELOCIDADE MEDIA DIARIA (AUT)(m/s) | na | rajada mínima do vento para o período | diário | 0 | na |
| na | nomeEstacao | nome que identifica a estação meteorológica | diário | 1 | varchar |
| na | statusEstacao | indicador se é operante [O] ou inoperante [I] | diário | 1 | char (1) |
| na | tipoEstacao | indicador se é automática (A) ou convencional [C] | diário | 1 | char (1) |

4.3 Análise do conjunto de dados de São Paulo / SP

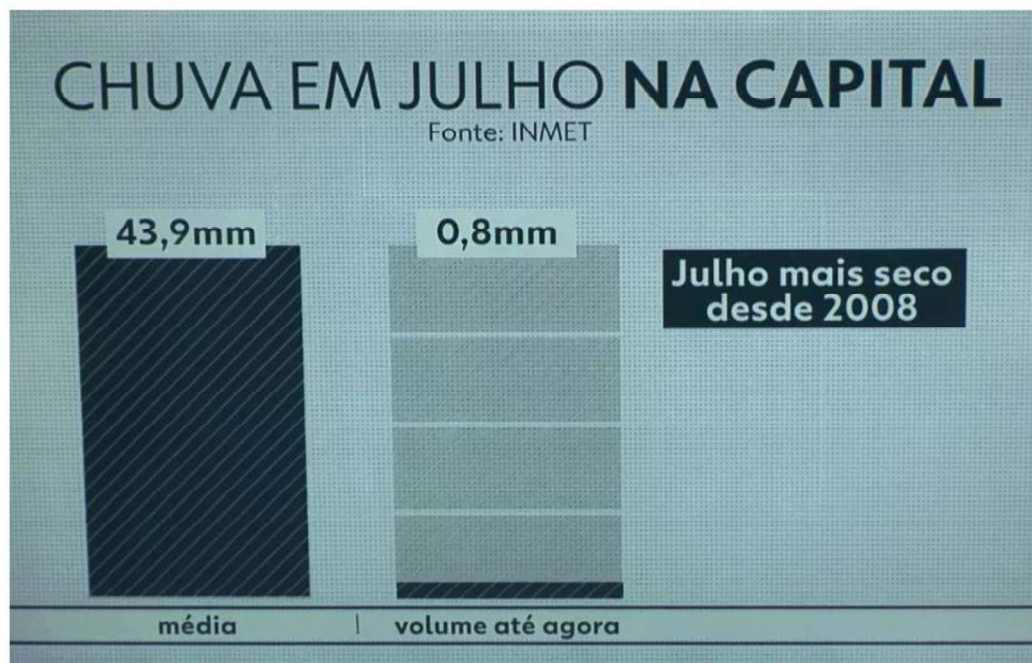
Adiante, pode-se verificar a construção de climogramas a partir dos dados obtidos pelo INMET. Neste exercício, foi utilizada, novamente, a estação automática do Mirante de Santana, por conter a maior completude dos registros dentre todas as estações apuradas.

A partir da leitura dos climogramas elaborados para a cidade de São Paulo entre 2008 e 2022, pode-se notar que, na média (figura 3), o desenho é similar ao que se tem em livros didáticos ou notícias de mídia que tratam o tema. Entretanto, são notórias as desigualdades e imprevisibilidades a cada ano, o que sugere que o senso comum de se basear pela média histórica pode levar a imprecisões, dado a variações em determinados meses e anos, como aponta o professor Antonio Carlos Zuffo, em entrevista para o G1: “...um sistema hídrico não pode considerar apenas os volumes de chuva médios, porque eles nunca ocorrem. E existem medidas de gestão para enfrentar situações atípicas” (2014).

Por exemplo, a figura 2 traz uma análise, realizada em 2017, sobre o tempo seco na cidade. Destaque para o valor de referência histórico, de 44 mm, o que é uma pluviosidade histórica similar ao calculado na figura 3, de 48 mm. Entretanto, é interessante destacar que os valores medidos para o mês de julho variam entre

valores inferiores a 10 mm (2008, 2011, 2016 e 2017) a valores superiores a 140 mm (2009 e 2019).

Figura 2: média histórica de pluviosidade para o mês de julho em São Paulo / SP. Fonte: g1.globo.com



SP tem julho mais seco desde 2008 — Foto: Reprodução/TV Globo

Outro ponto de destaque diz respeito ao período que antecede a chamada “crise hídrica de 2014”, cuja representação gráfica pode ser verificada na figura 5. A média calculada para o período de 2008 a 2022 é de 226 mm para o mês de dezembro (vide figura 3). Em 2013, conforme figura 4, esse valor foi de apenas 80 mm, ou 35% do esperado para o mês. Já em 2022 (figura 6), a medição foi de 166 mm em dezembro.

Segundo a SABESP, “Foi a pior seca desde que as medições dos institutos de meteorologia começaram, há 84 anos. Em dezembro de 2013, foi registrado índice 72% menor que a média. Em janeiro e fevereiro choveu cerca de 65% menos que o normal” (2014). Contudo, de acordo com os dados que foram calculados neste trabalho (leva-se em conta que os períodos históricos são diferentes, o que pode levar a médias históricas diferentes também), é verdade que em dezembro/2013 houve uma redução na variação, 72% segundo a SABESP e 65% neste trabalho, com um desvio, provavelmente, determinado pela diferença dos períodos de médias históricas. Mas para janeiro e fevereiro/2014 há uma grande diferença nas apurações nos valores de referência, 65% para a SABESP e 23% neste trabalho, o que a diferença nos períodos

não seria suficiente para justificar a mudança de escala dos números. Embora o foco deste estudo não seja a verificação das declarações de mídia, este pode ser um dos usos dos dados meteorológicos estruturados, bem armazenados e com indicadores e gráficos adequados. Por fim, fica uma reflexão: como os dados já apontavam uma queda abrupta na pluviosidade em dezembro/2013, os monitoramentos e avisos à população poderiam ser feitos com mais antecedência, mitigando os impactos de 2014.

Figura 3: climograma da cidade de São Paulo (Mirante de Santana) entre 2008 e 2022. Fonte: elaborada pelo autor

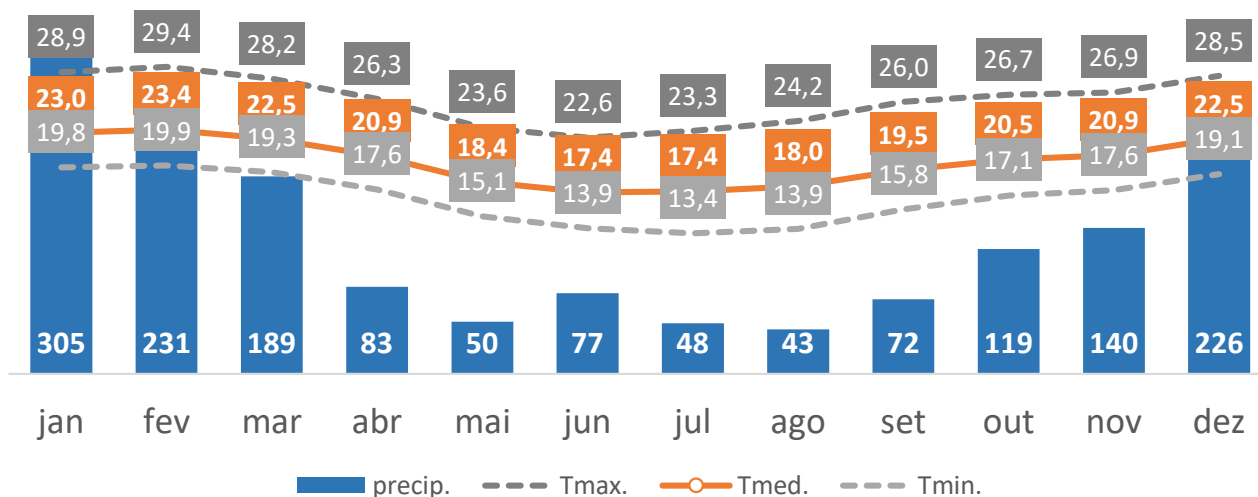


Figura 4: climograma da cidade de São Paulo (Mirante de Santana) em 2013. Fonte: elaborada pelo autor

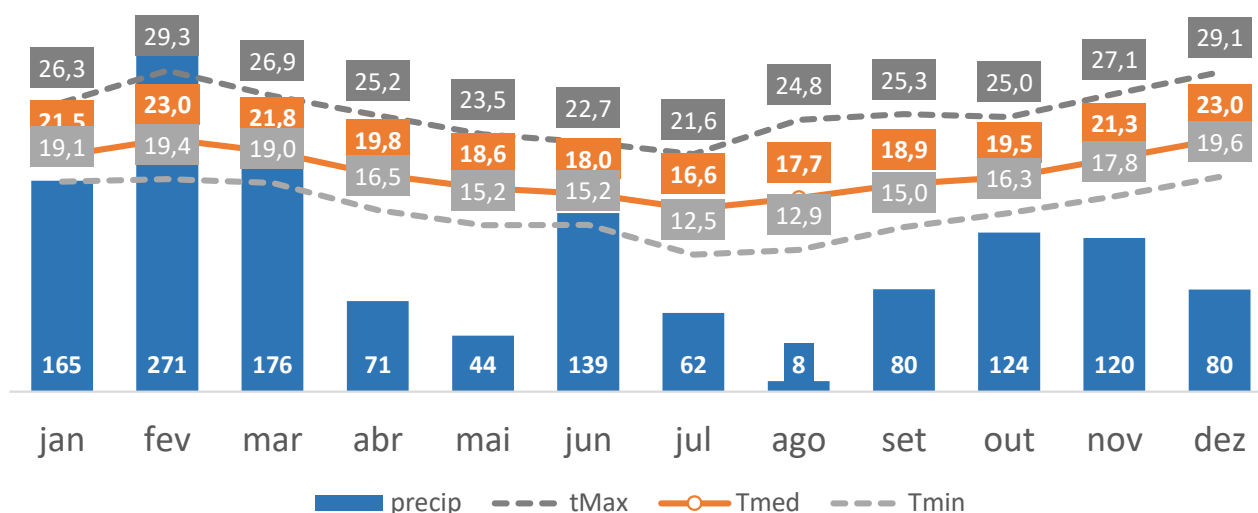


Figura 5: climograma da cidade de São Paulo (Mirante de Santana) em 2014 Fonte: elaborada pelo autor

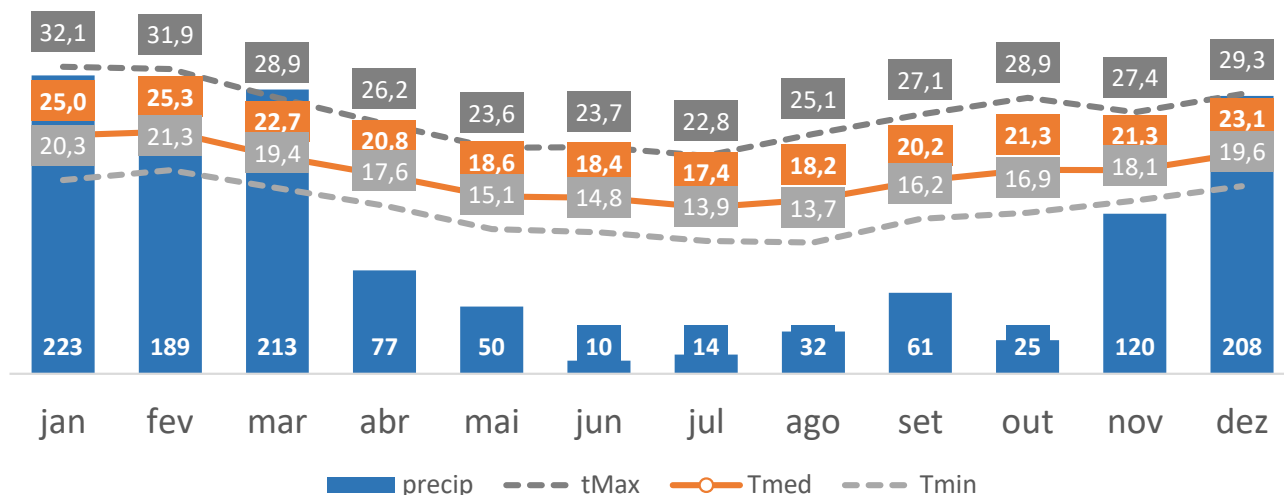
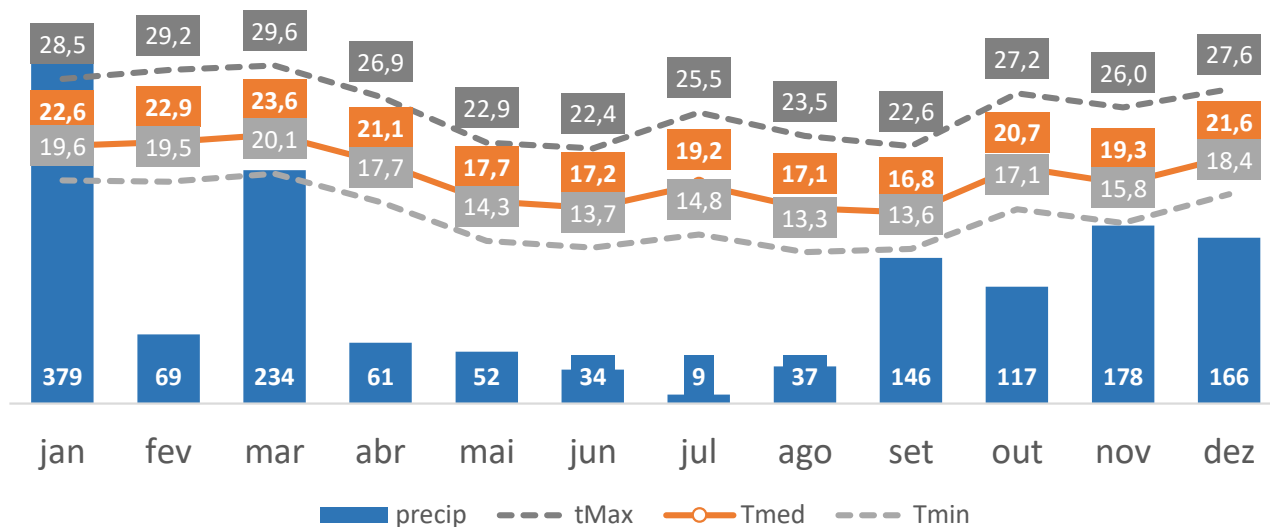


Figura 6: climograma da cidade de São Paulo (Mirante de Santana) em 2022 Fonte: elaborada pelo autor



4.4 Análise do conjunto de dados de São Luiz do Paraitinga / SP

Como pode ser verificado na figura 7, a estação automática de São Luiz do Paraitinga operava com relativa confiabilidade até o ano de 2018. As razões para uma menor completude das informações a partir de 2019 não fazem parte do escopo deste trabalho, embora seja uma ideia de contribuição futura. Deste modo, o período de

análise compreendeu de 2008 a 2018, cujo intervalo abarcou o evento de chuvas no verão de 2009/2010.

O município de São Luiz do Paraitinga ficou conhecido por abrigar uma grande enchente na virada do ano de 2009 para 2010. Como aponta Moradei (2016, p.109), “De fato, 2009 foi um ano atípico. Choveu muito durante todo o ano. Houve transbordamento do Rio Paraitinga em julho, período, normalmente, de seca”. Julho de 2009 foi um ano em que a pluviosidade apontou para 131 mm (figura 9), enquanto a média do período (2008 a 2018, figura 7) revela 39 mm. Ou seja, uma variação de 236% em relação a uma normalidade calculada a partir dos dados do INMET.

Ao analisar os dados de 2009, pode-se concluir que todos os meses a partir de junho foram superiores à média calculada do período de 2008 a 2018. Essa diferença teve seu ápice em dezembro/2009 (figura 8), quando houve 396 mm de chuvas contra 210 mm de média (figura 7), configurando um aumento de 186 mm, quase o dobro. Entretanto, um monitoramento mais persistente já revelaria o potencial da cheia do Rio Paraitinga, que levou à destruição de parte do município.

Figura 7: climograma da cidade de São Luiz do Paraitinga – 2008 a 2018. Fonte: elaborada pelo autor

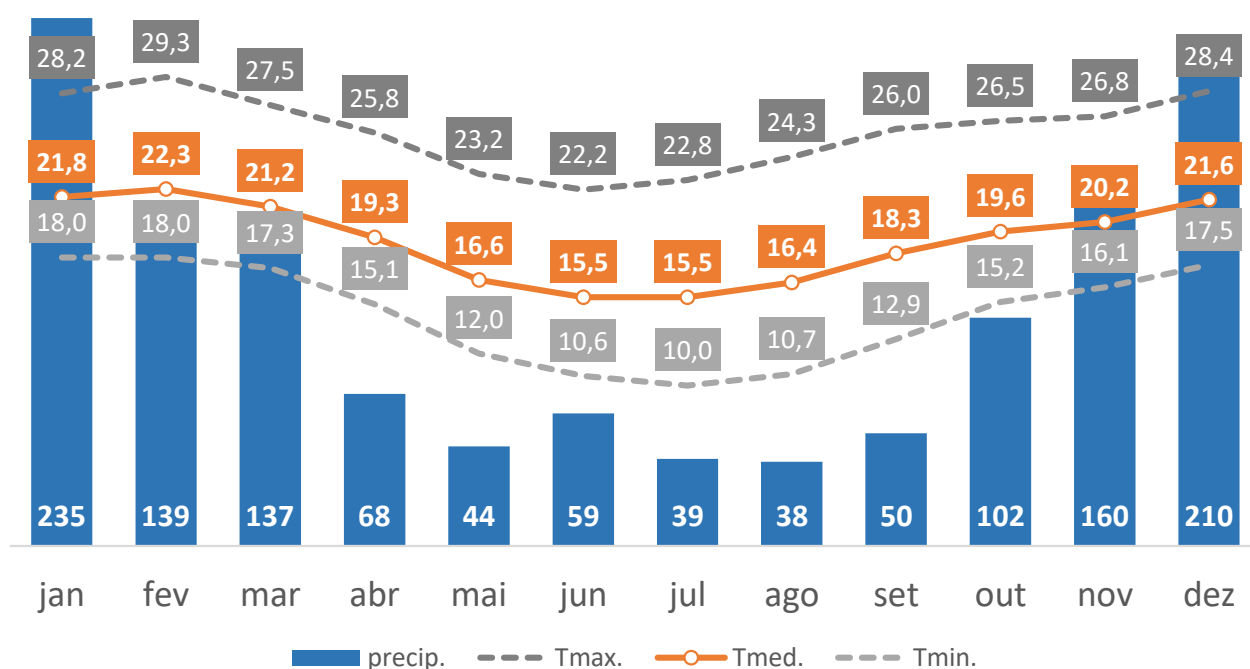


Figura 8: climograma da cidade de São Luiz do Paraitinga – 2009. Fonte: elaborada pelo autor

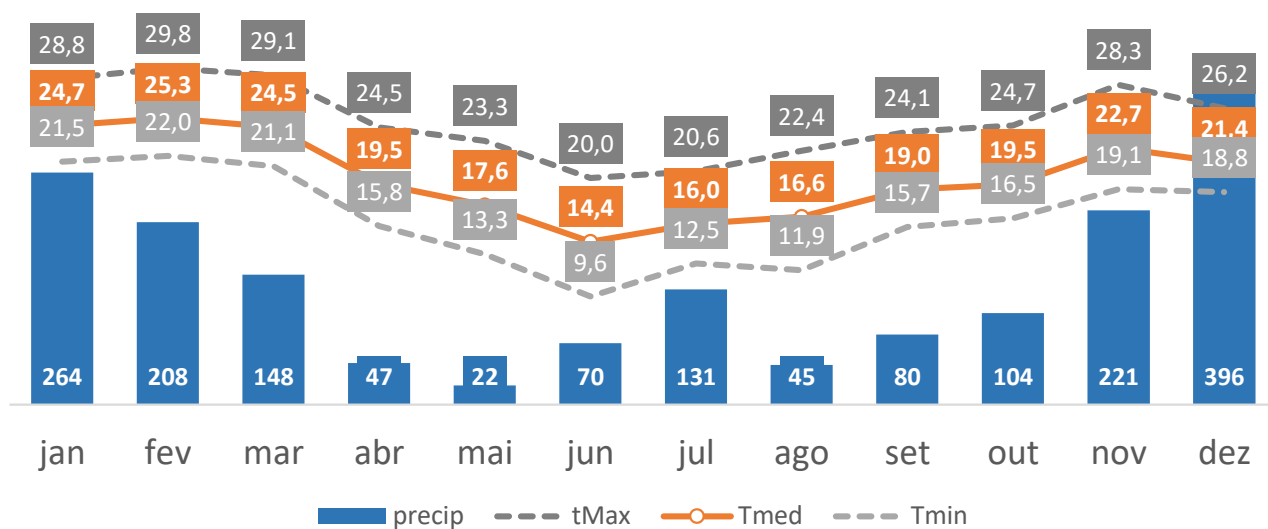
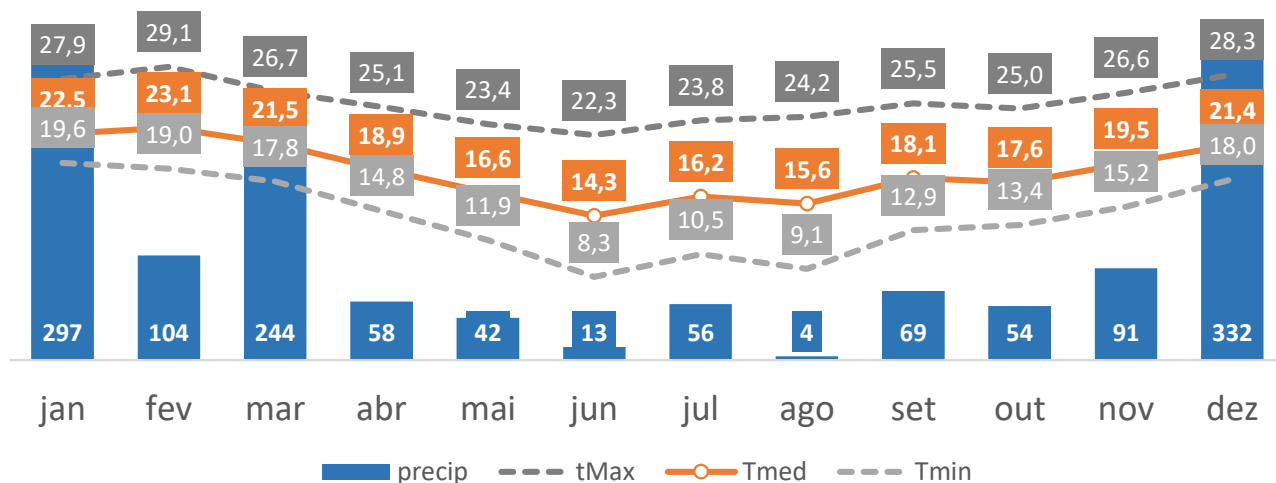


Figura 9: climograma da cidade de São Luiz do Paraitinga – 2010. Fonte: elaborada pelo autor



5 Conclusão

Este trabalho espera contribuir com informações que apoiem estudantes, pesquisadores e qualquer pessoa interessada em analisar as informações meteorológicas disponíveis a partir da aplicação do modelo de ciclo de vida proposto. Não houve proposta de criar uma solução definitiva para questões levantadas ao longo das análises, mas, sim, de facilitar o acesso e compreensão dos dados meteorológicos disponíveis.

Como visto anteriormente, uma análise pautada no modelo de Ciclo de Vida permite a organização do processo da geração da informação como um todo e, assim, gera melhores possibilidades de continuidade dos trabalhos. Assim sendo, as respostas se tornam mais eficazes e a informação percorre um caminho mais fluido.

A aplicação do modelo de ciclo de vida dos dados – DataONE se revelou bastante satisfatória, onde o método pode ser aplicado a variados assuntos. Espera-se, dessa maneira, que os trabalhos que envolvem análise de dados dessa natureza se tornem cada vez mais fluidos, organizados, replicáveis e documentados.

5.1 Contribuições do trabalho

Como contribuições, destaca-se a possibilidade de compreensão de massas de dados climatológicos, como, por exemplo, se valendo de climogramas a partir de recortes específicos de períodos, além de geração de informações mais atualizadas.

Outro ponto é a possibilidade de replicação do trabalho a partir do modelo de gestão de dados, contribuindo com a comunidade científica, evitando possíveis retrabalhos. Do ponto de vista específico sobre o tema da climatologia, o modelo pode ser aproveitado para nova construção de climogramas e demais análises visuais que facilitam o entendimento, gerando informações mais atualizadas, precisas e confiáveis. Em um contexto de mudanças abruptas no meio ambiente, é recomendável que as análises também acompanhem esse movimento.

O apoio de um modelo de gestão do ciclo de vida de dados para monitoramento mais tempestivo e organizado por veículos de mídia, tornando a comunicação mais efetiva, é algo a ser considerado em virtude da necessidade de notícias que esclarecem eventos climáticos, produzindo e fornecendo informações em linguagem mais acessível.

5.2 Trabalhos futuros

A confrontação de dados pluviométricos e de temperatura com informações complementares (por exemplo: saúde, doenças sazonais, desmatamento) são objetivos futuros, tendo em vista a continuidade do trabalho e a necessidade de respostas cada vez mais acuradas em torno do tema das mudanças climáticas.

Outro tema de contribuição é o confronto dos dados climatológicos com a organização política do meio ambiente, como o aumento ou redução de temperaturas em regiões que constituem extensas áreas de espaços protegidos em relação a outros espaços que estão degradados, ou se há correlação entre empreendimentos de significativo impacto ambiental com mudanças de patamar nos dados coletados. Ou, ainda, se houve alguma mudança no regime de chuvas ou no padrão de temperatura após a criação de áreas de proteção de determinada região.

Adicionalmente, outras formas de gestão de dados devem ser testadas, com a finalidade em se encontrarem modelos cada vez mais adequados no processo de coleta, armazenamento, documentação e análise de dados climatológicos.

Referências Bibliográficas

INMET. **Dados Meteorológicos. Banco de Dados Meteorológicos.** Disponível em: <https://bdmep.inmet.gov.br/>. Acesso em 27 abr. 2023

ANJOS, R. L. da.; DIAS, G. A. **Atuação dos Profissionais da Informação no Ciclo de Vida dos Dados. Um Estudo Comparado.** Revista do Programa de Pós-Graduação em Ciência da Informação da Universidade Estadual de Londrina (2019): 80-101. DOI: 10.5433/1981-8920.2019v24n1p80. Acesso em: 05 dez. 2023

SILVA, J. A. F; CARPENEDO, C. B.; SILVA, C. B. da. **Precipitação Pluvial Mensal Em Uberlândia - MG Obtida Por Diferentes Fontes De Dados.** Revista Brasileira De Climatologia 31 (2022): 26-48. <https://doi.org/10.55761/abclima.v31i18.15846> Acesso em: 08 ago. 2023

SABINO, M.; SOUZA, A. **Gap-filling Meteorological Data Series Using the GapMET Software in the State of Mato Grosso, Brazil.** Revista Brasileira De Engenharia Agrícola E Ambiental 27.2 (2023): 149-56. <https://doi.org/10.1590/1807-1929/agriambi.v27n2p149-156>. Acesso em: 10 jul. 2023

SILVA, F. C. C. da. **Visualização de dados: passado, presente e futuro.** Liinc em Revista, [S. l.], v. 15, n. 2, 2019. DOI: 10.18617/liinc.v15i2.4812. Disponível em: <https://revista.ibict.br/liinc/article/view/4812>. Acesso em: 10 jul. 2023.

BIER, A. A.; FERRAZ, S. E. T. **Comparação de Metodologias de Preenchimento de Falhas em Dados Meteorológicos para Estações no Sul do Brasil (2015).** Revista Brasileira de Meteorologia, v. 32, n. 2, 215-226, 2017. <http://dx.doi.org/10.1590/0102-77863220008>

MORADEI, N. S. **A grande enchente de São Luiz do Paraitinga – 2010 –** São Paulo, 2016. <https://doi.org/10.11606/D.16.2017.tde-02032017-113404>. Acesso em 10 jul. 2023

G1 SP. **Há 45 dias sem chuva, SP tem julho mais seco desde 2008, diz Inmet.** 2017. Disponível em: <https://g1.globo.com/sao-paulo/noticia/ha-45-dias-sem-chuva-sp-tem-julho-mais-seco-desde-2008-diz-inmet.ghtml>. Acesso em 30 out. 2023

G1 SP. **A chuva já resolveu a crise da água em SP? Veja perguntas e respostas.** 2014. Disponível em: <https://g1.globo.com/sao-paulo/noticia/2014/11/chuva-ja-resolveu-crise-da-agua-em-sp-veja-perguntas-e-respostas.html>. Acesso em 30 out. 2023

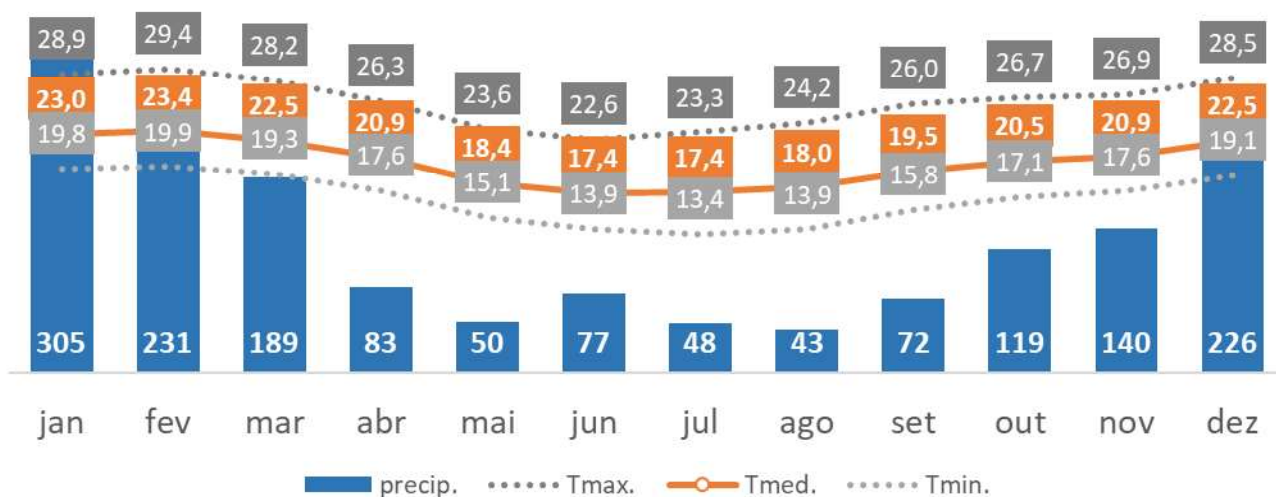
Jornal da Reconstrução. **Como tudo aconteceu.** Disponível em: https://www.saoluizdoparaitinga.sp.gov.br/site/wp-content/uploads/2012/02/jr_n1.pdf. Acesso em 04 nov.2023

ZULLO JUNIOR, J.; PINTO, H. S.; PINTO, D. S. **Organização e consistência de banco de dados pluviométricos diários.** Disponível em: <https://www.sbagro.org/files/biblioteca/1848.pdf>. Acesso em 04/11/2023

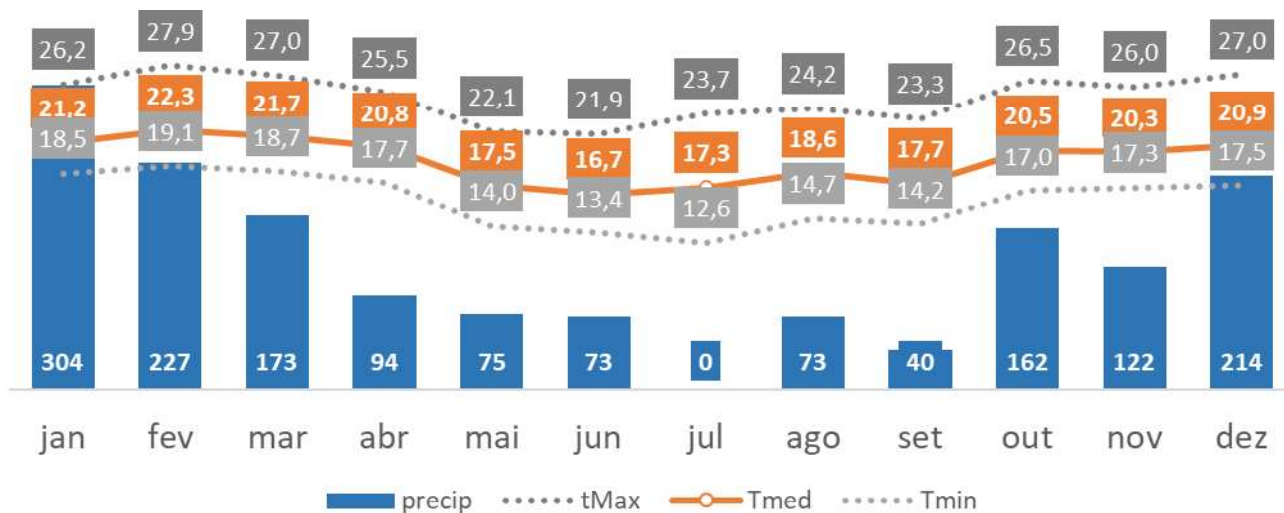
Apêndice: Climogramas de São Paulo / SP

Esta seção se refere à construção de climogramas do município de São Paulo / SP no período compreendido entre os anos de 2008 e 2022.

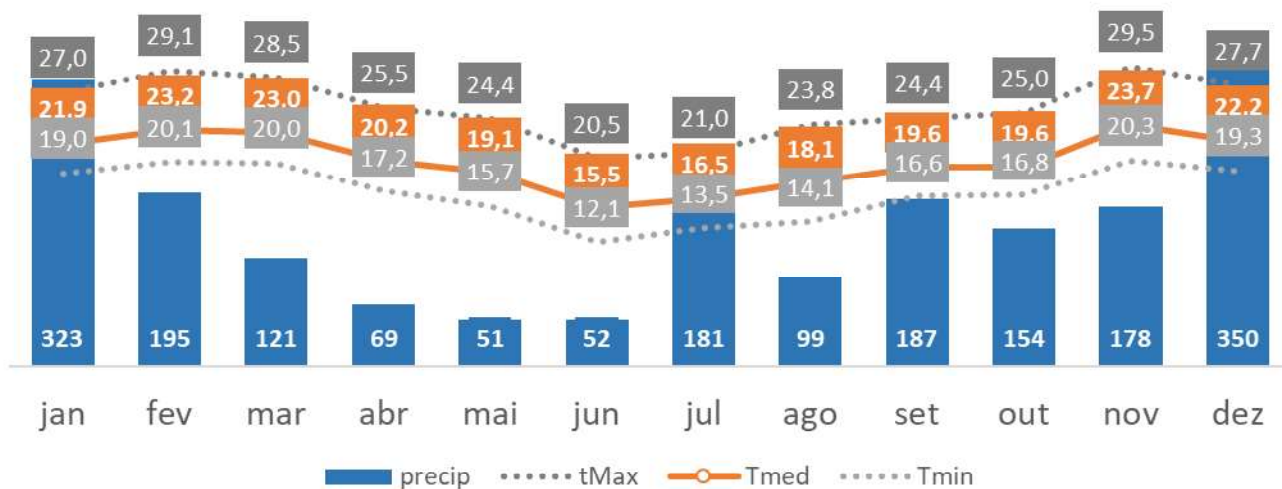
1) 2008 a 2022 (consolidado)



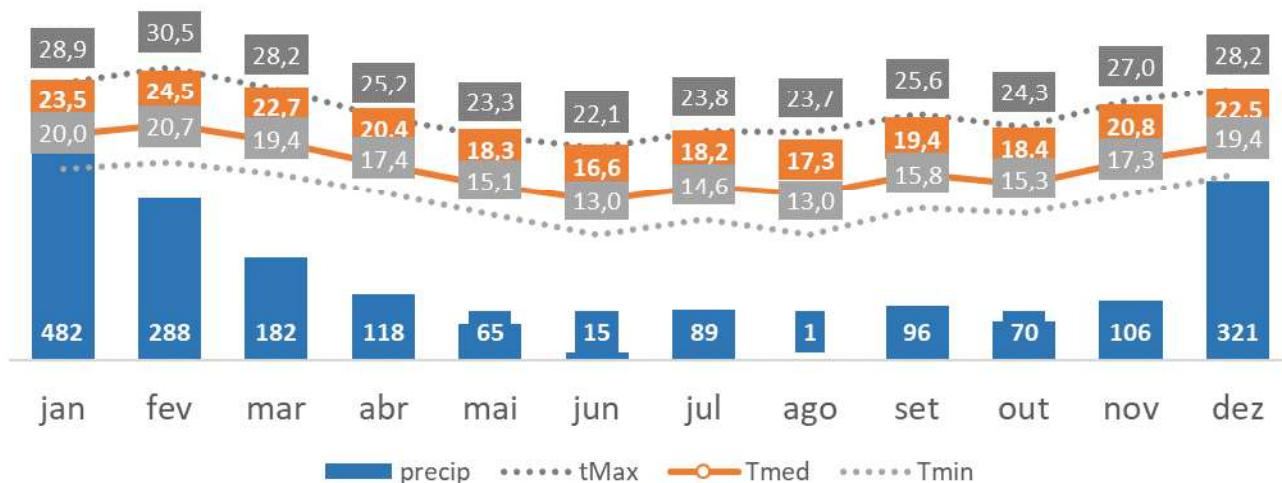
2) 2008



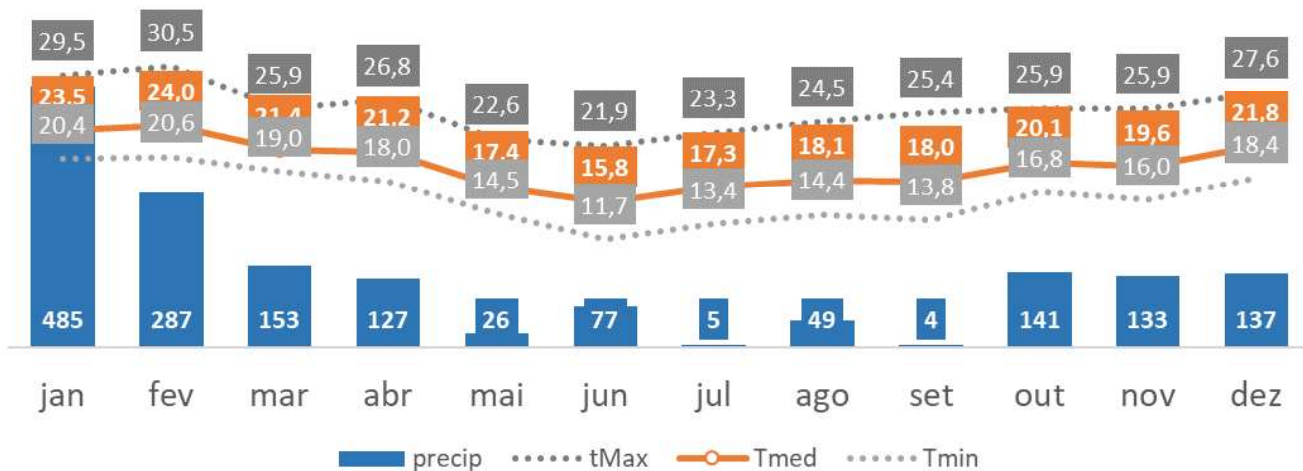
3) 2009



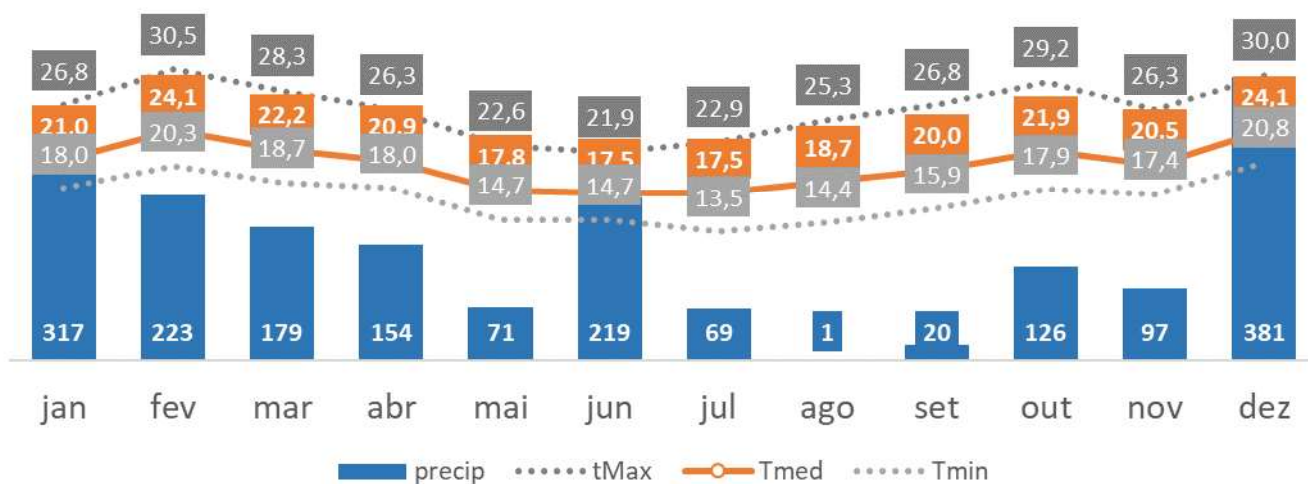
4) 2010



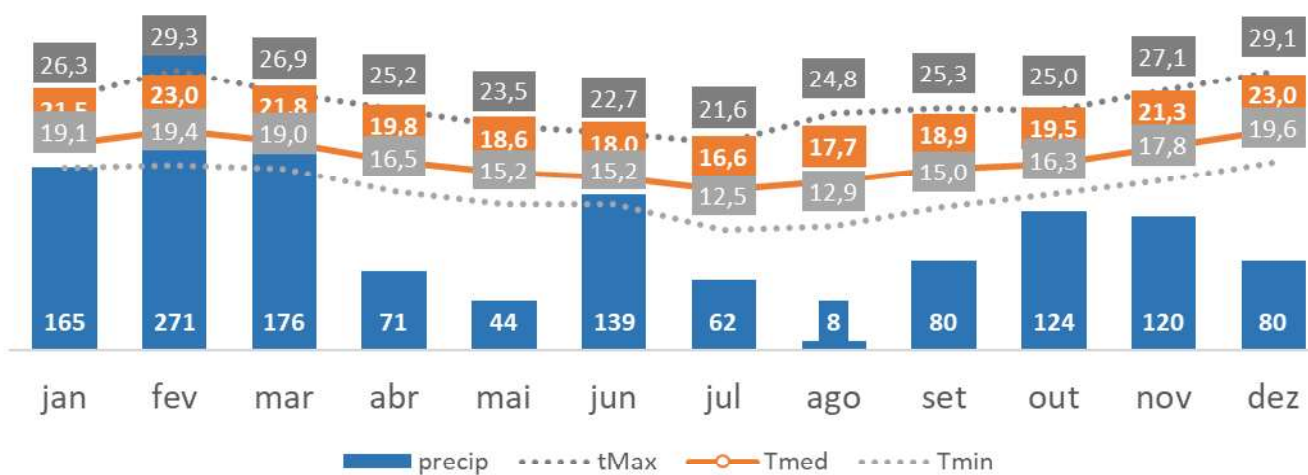
5) 2011



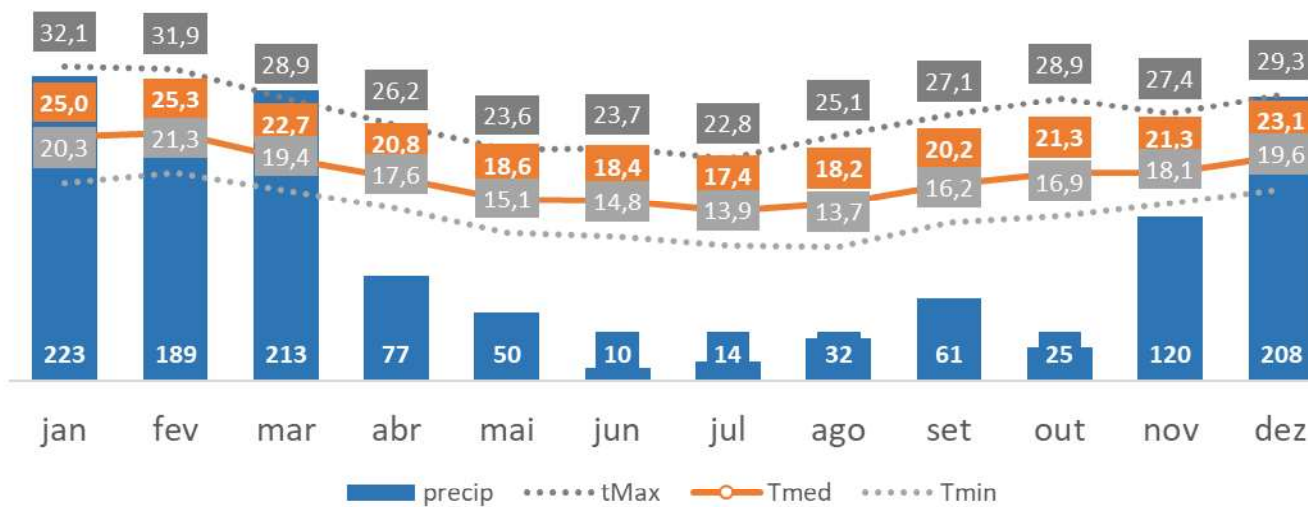
6) 2012



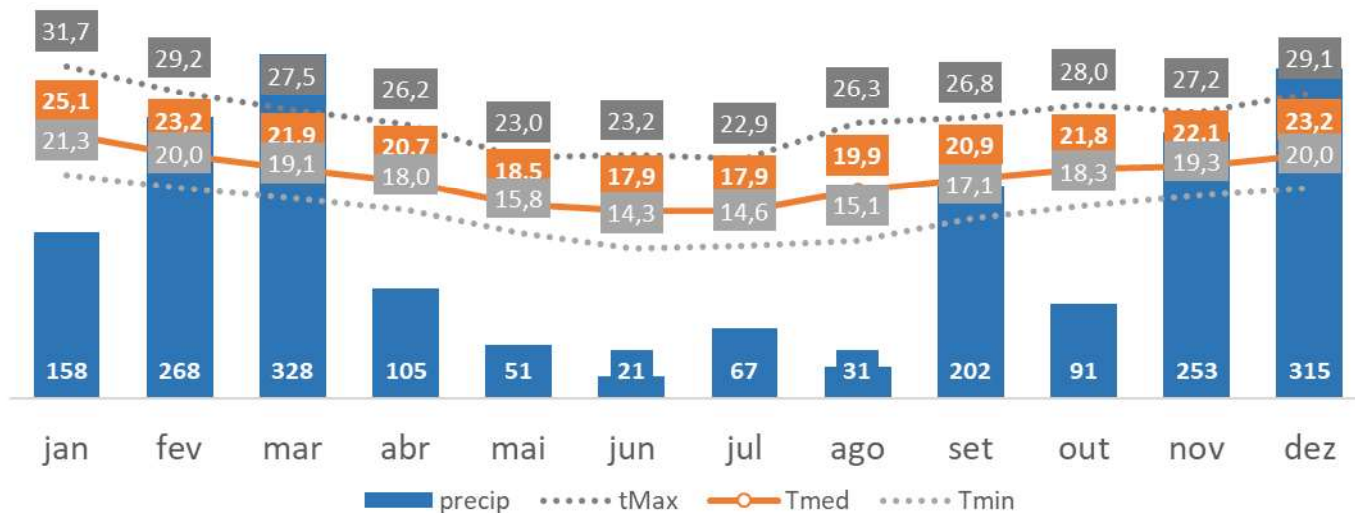
7) 2013



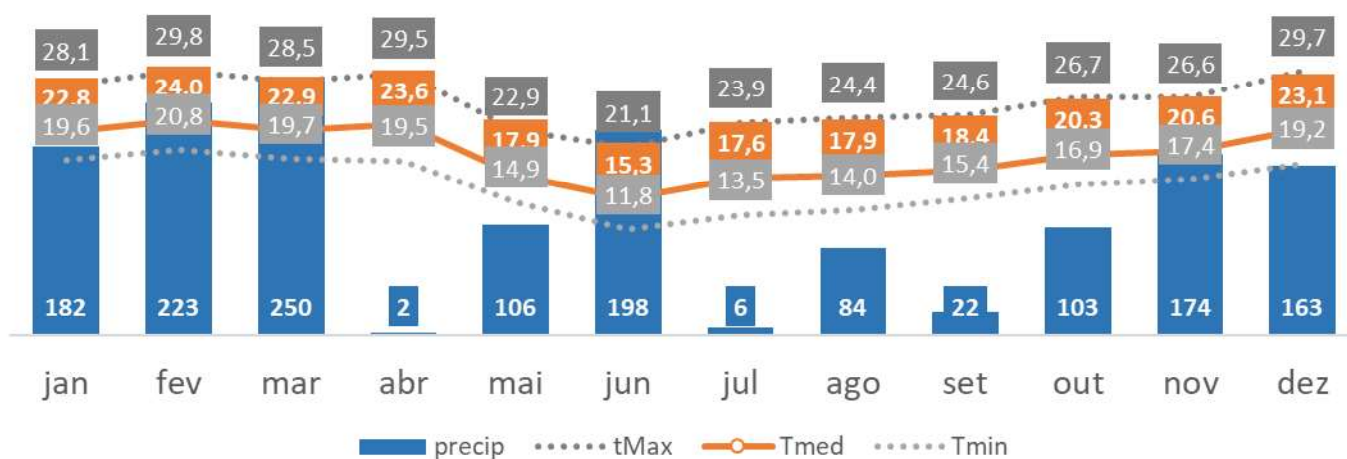
8) 2014



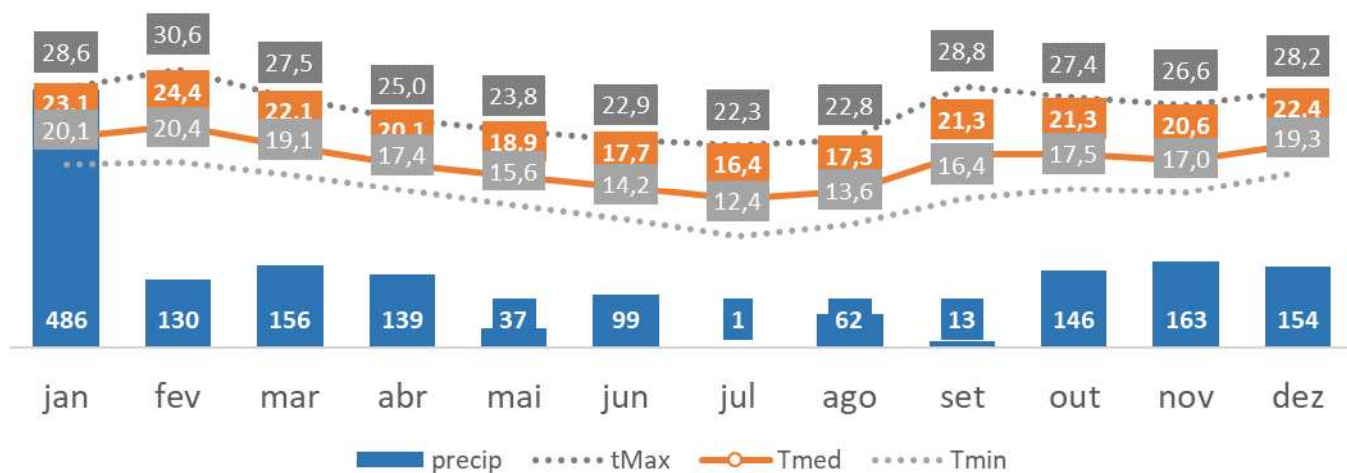
9) 2015



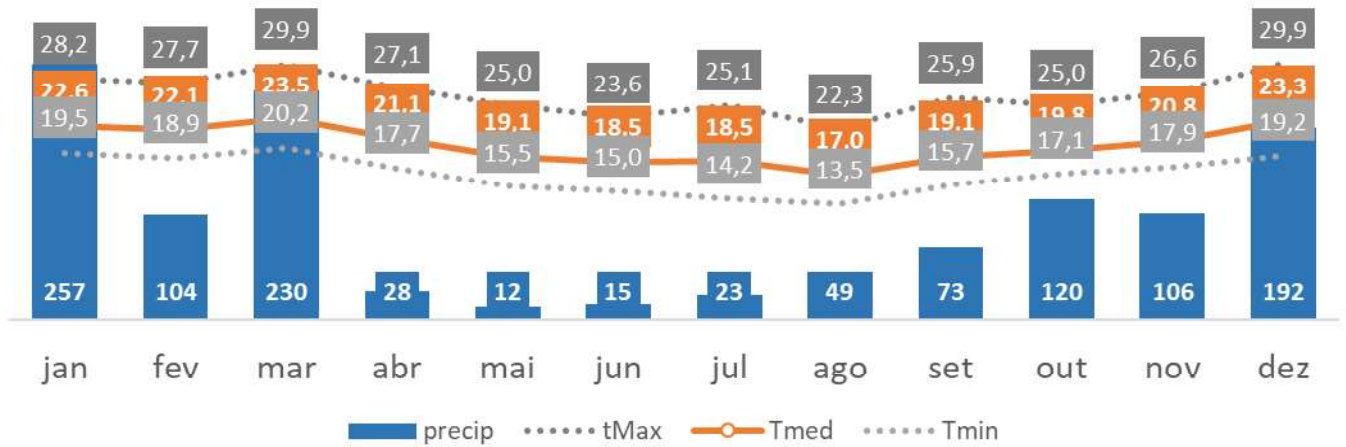
10) 2016



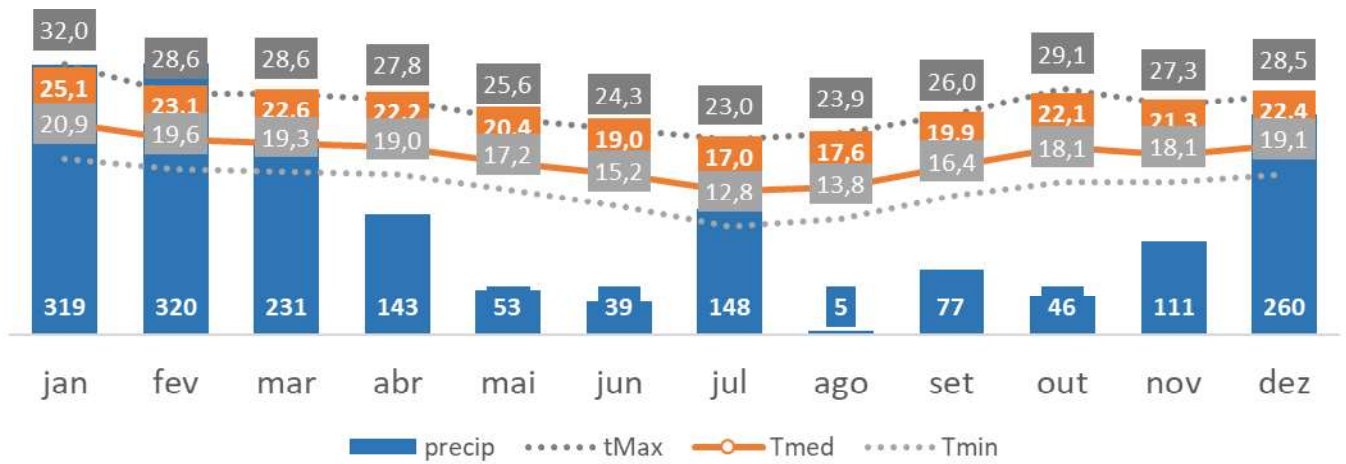
11) 2017



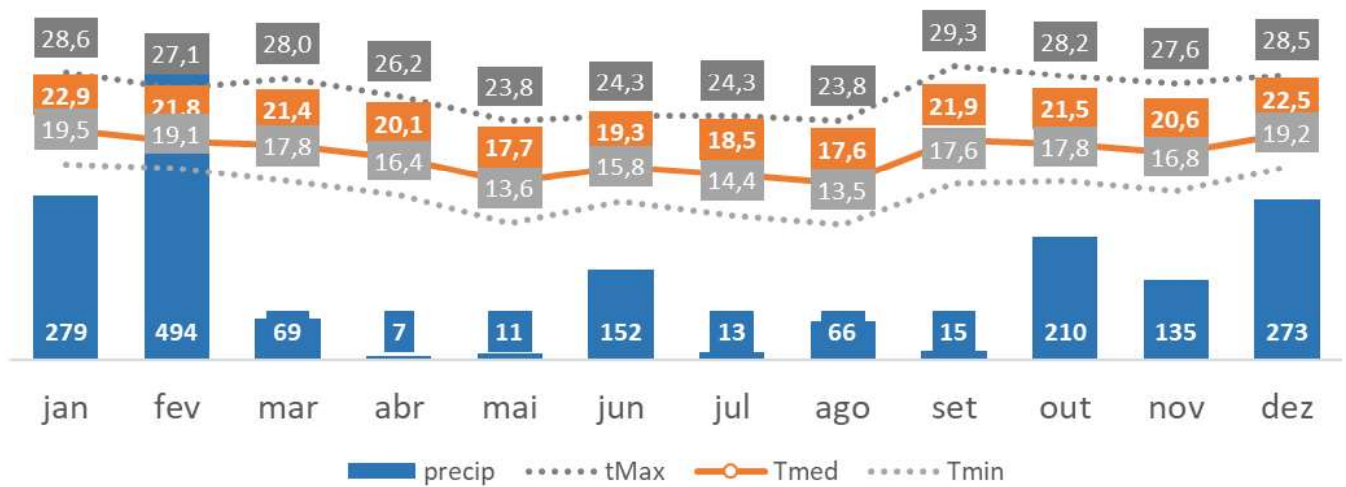
12) 2018



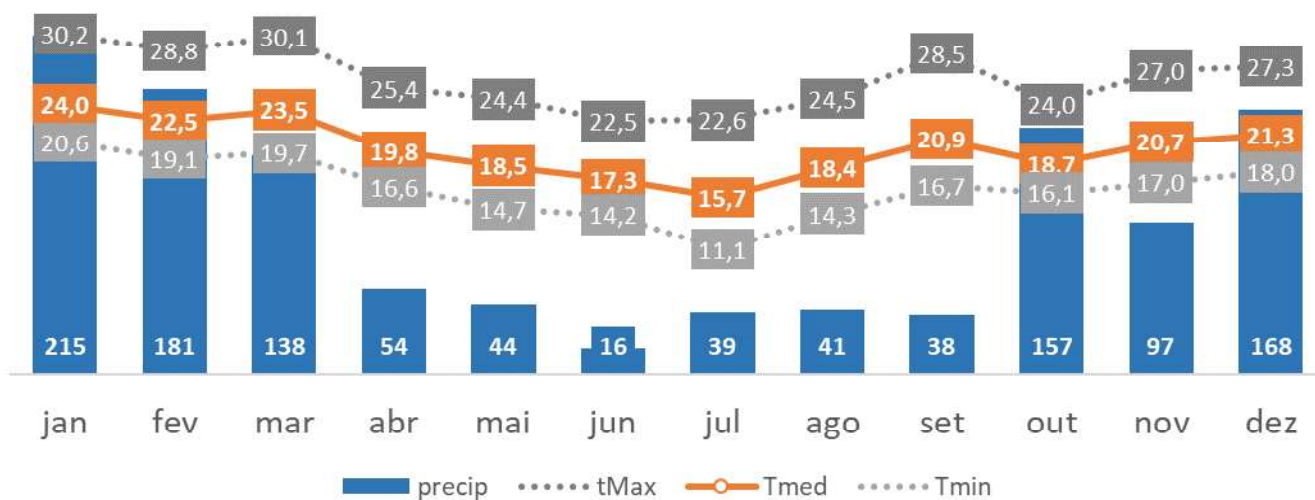
13) 2019



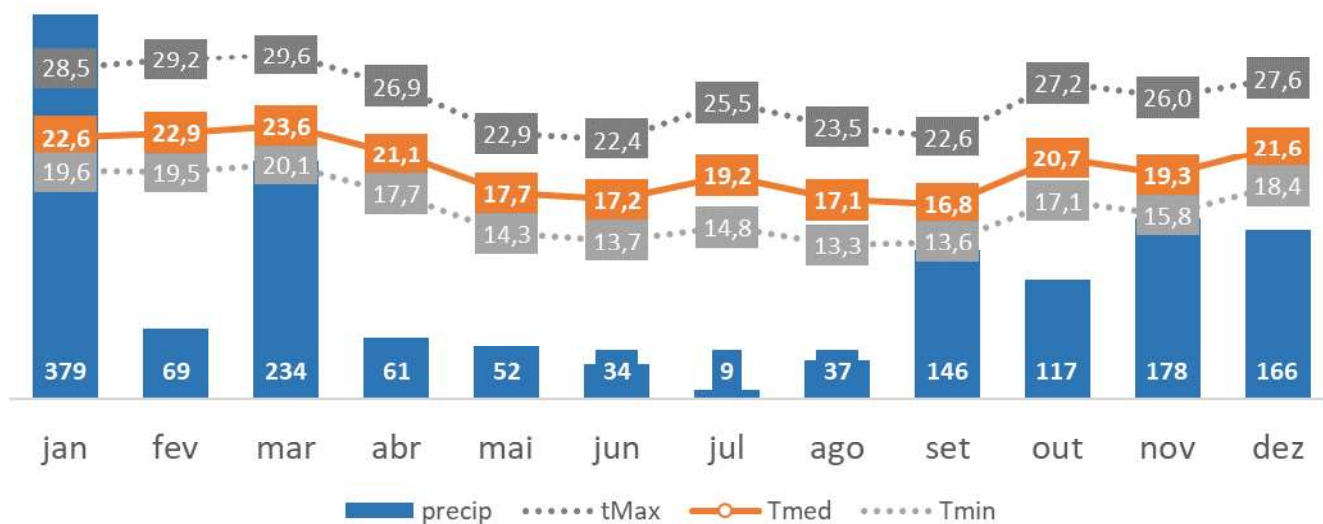
14) 2020



15) 2021



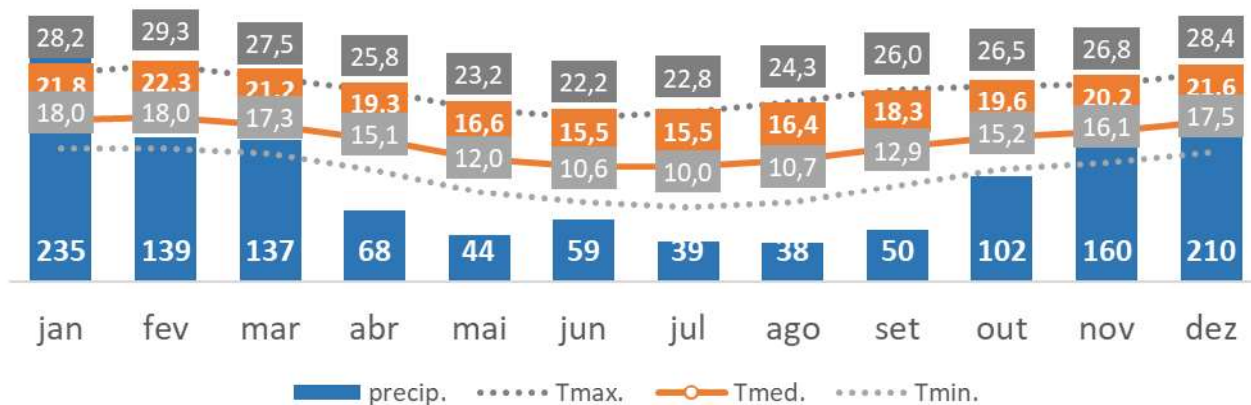
16) 2022



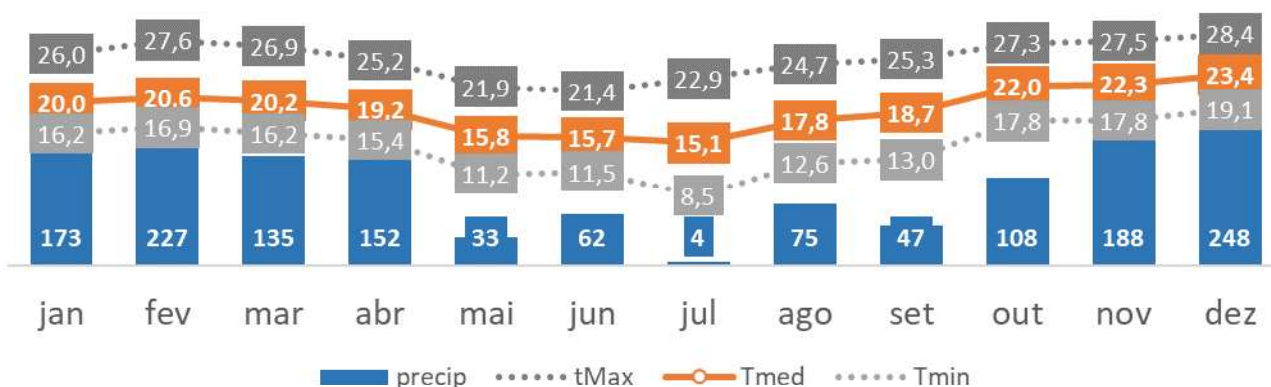
Apêndice: Climogramas de São Luiz do Paraitinga / SP

Esta seção se refere à construção de climogramas do município de São Luiz do Paraitinga / SP no período compreendido entre os anos de 2008 e 2018.

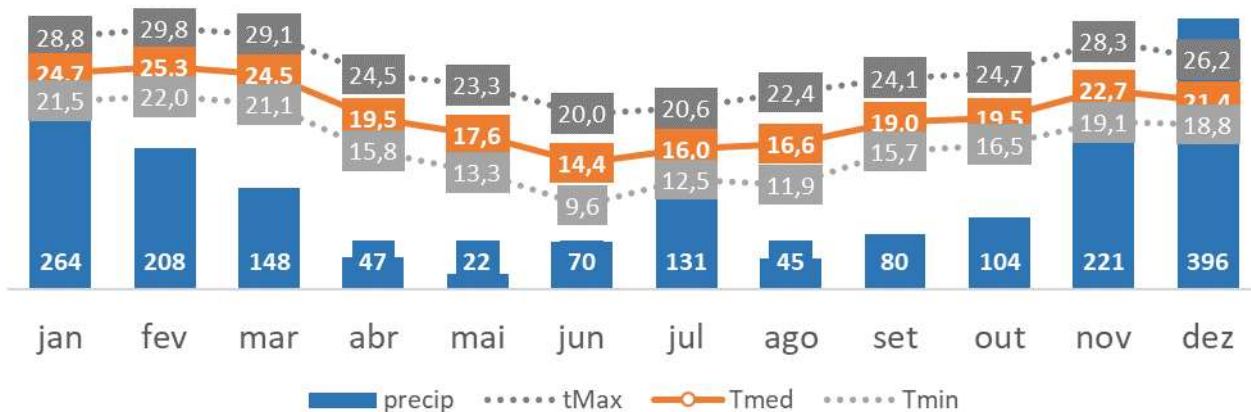
1) 2008 a 2018 (consolidado)



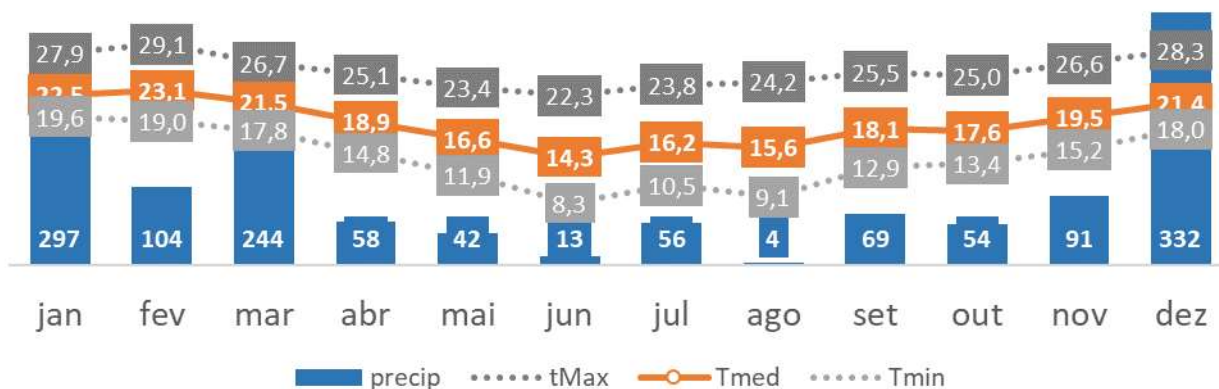
2) 2008



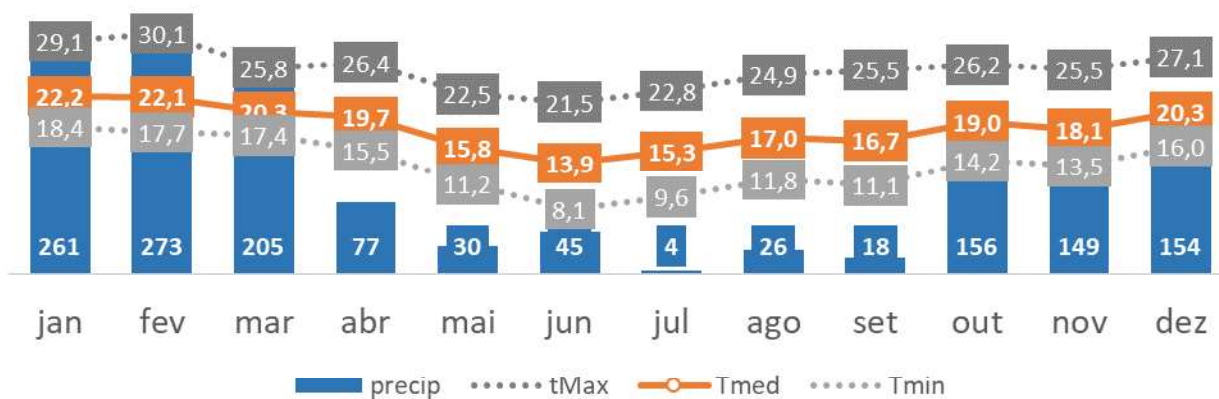
3) 2009



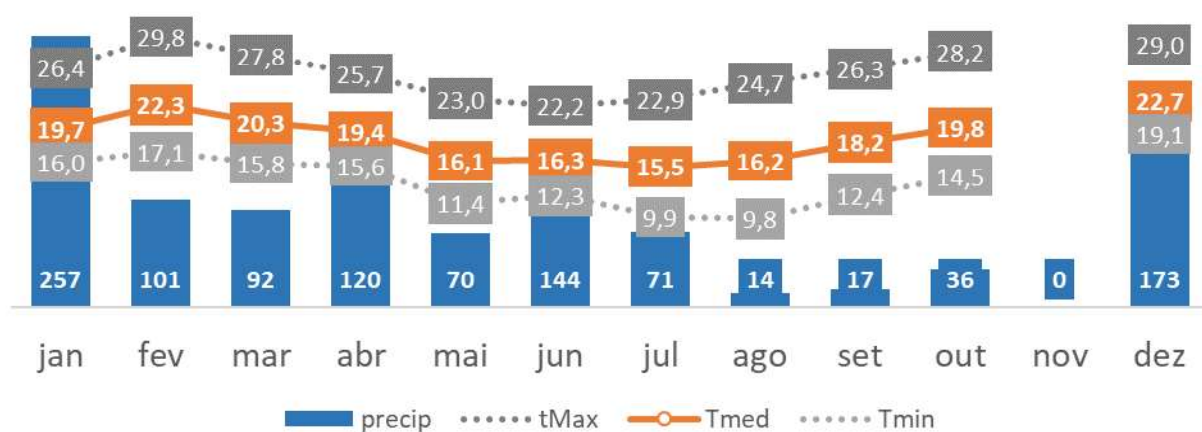
4) 2010



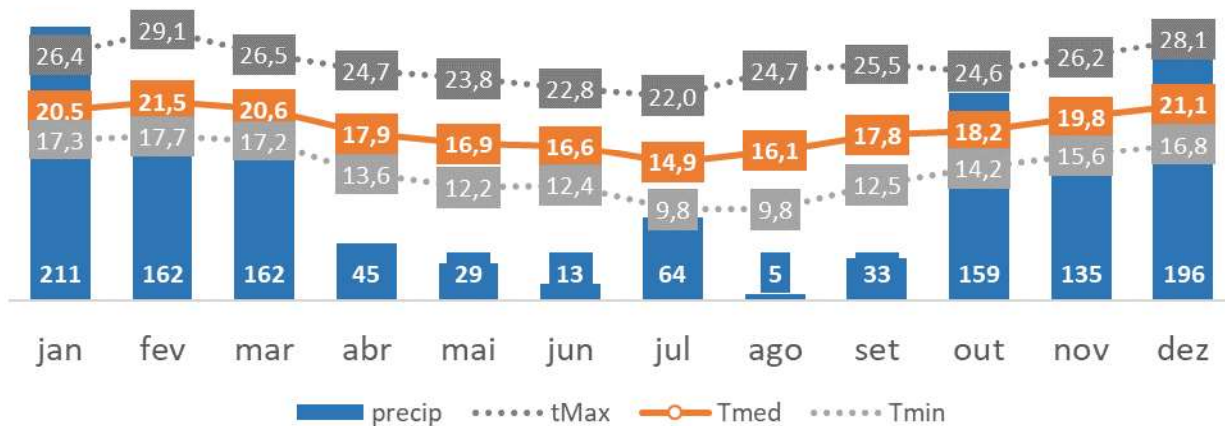
5) 2011



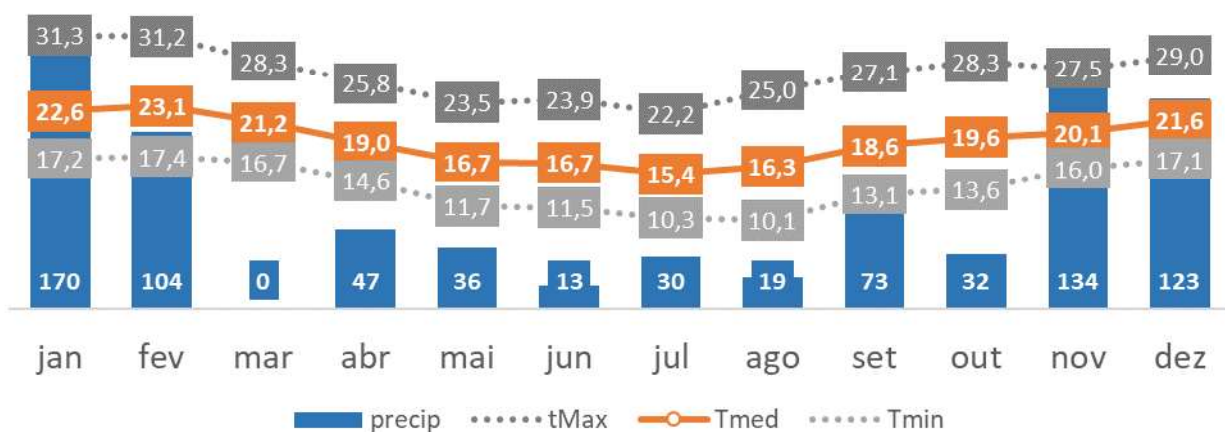
6) 2012



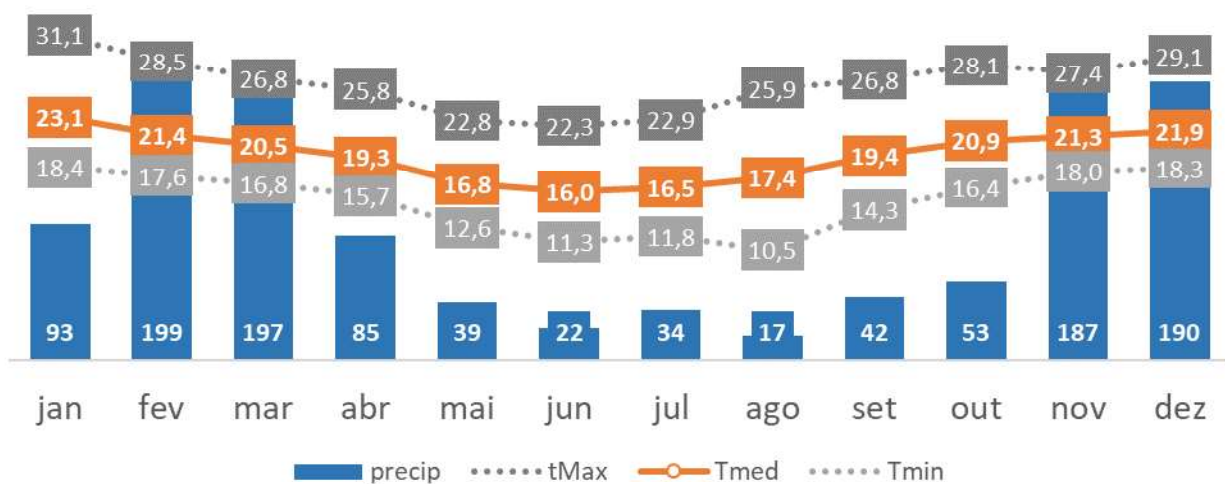
7) 2013



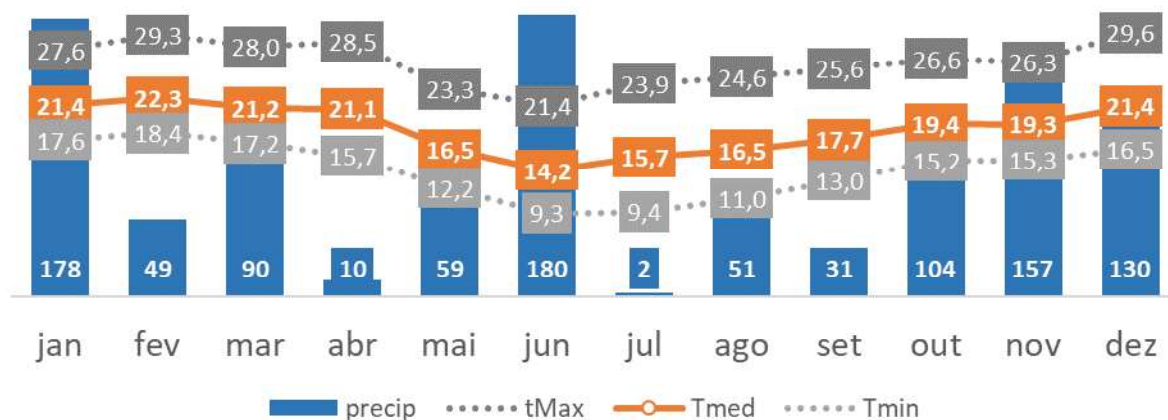
8) 2014



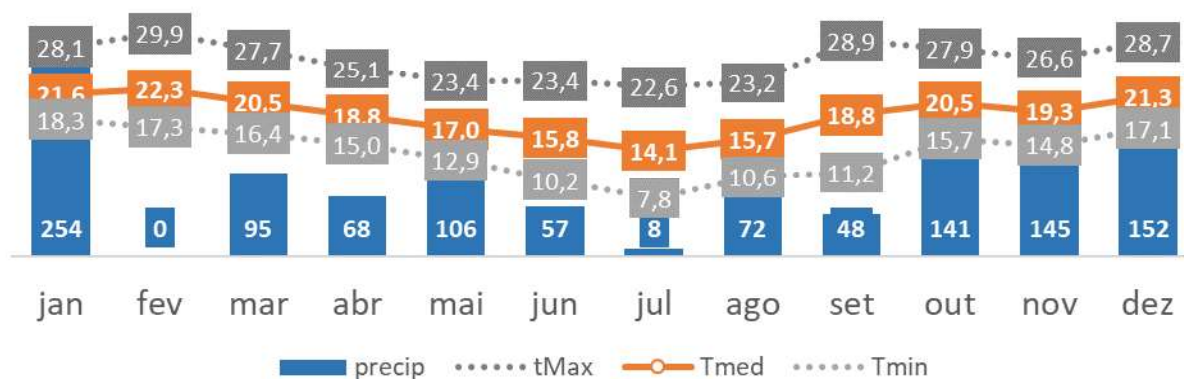
9) 2015



10) 2016



11) 2017



12) 2018

