

Desemprego no Brasil

Fernando Hannaka - 2 de dezembro de 2020

Introdução

Trabalho de conclusão do curso de Visualização da Informação da FGV-EMAp. O objetivo deste trabalho é elaborar uma visualização para obter insights da situação da força de trabalho no Brasil, com foco na taxa de desemprego.

Aquisição de dados

Sobre a situação de trabalho no Brasil, a Pesquisa Nacional por Amostra de Domicílios Contínua - PNAD Contínua do IBGE fornece dados desde 2012. A taxa de desemprego refere-se a desocupação, isto é, ao percentual de pessoas ativas na força de trabalho que não estão empregadas por pelo menos 40h semanais. Pessoas que não trabalham mas que não estão a procura de emprego não são consideradas no cálculo (<https://www.ibge.gov.br/explica/desemprego.php> e https://editalconcursosbrasil.com.br/blog/economia_desemprego-no-brasil/).

Pesquisa

O site do IBGE fornece algumas visualizações dos dados de taxa de desemprego porém não é possível realizar a comparação pelo tempo e por estados/regiões. Seguem exemplos abaixo:

População brasileira, de acordo com as divisões do mercado de trabalho, 2º trimestre 2020

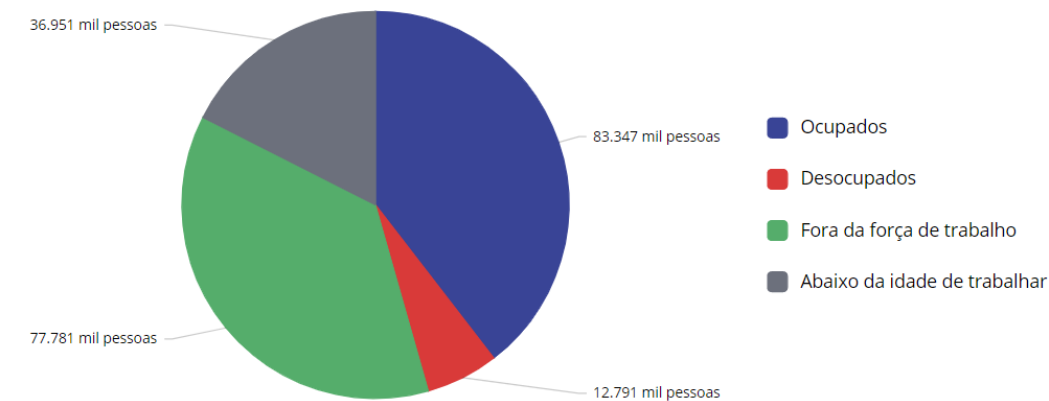


Figura 1 - Gráfico de pizza



Figura 2 - Gráfico de linha

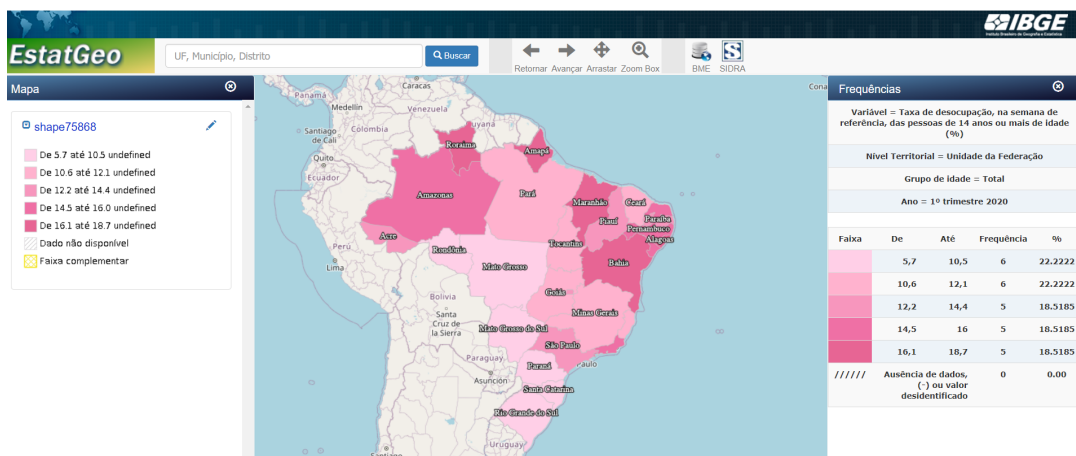
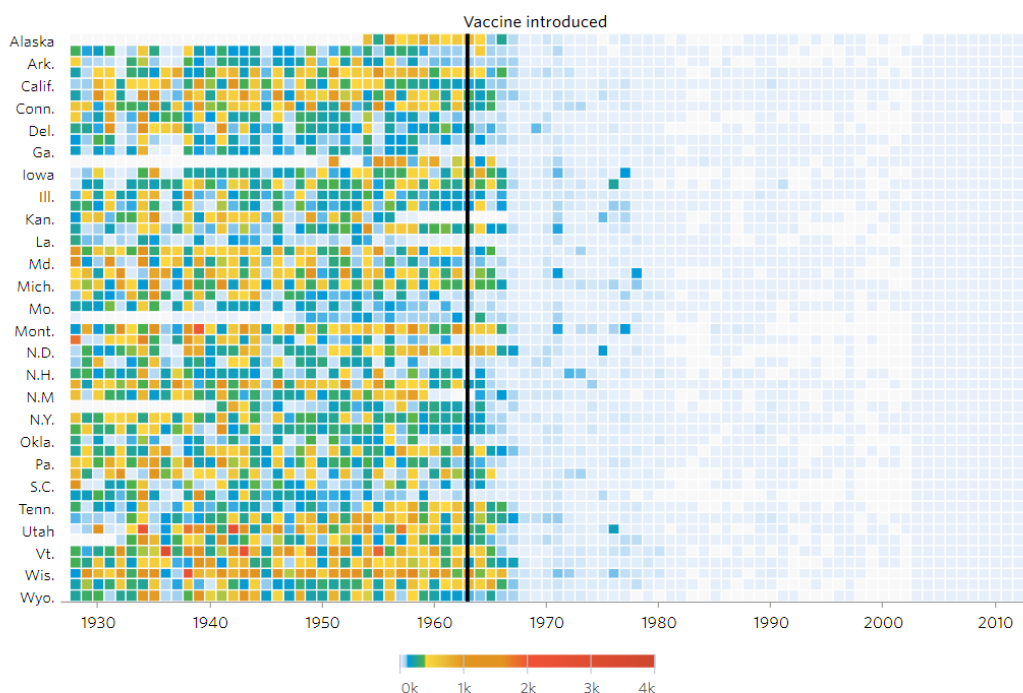


Figura 3 - Cartograma

A inspiração para a visualização que desejo fazer é o mapa de calor criado pelo The Wall Street Journal que mostra o impacto da vacina para doenças infecciosas no século 20, http://graphics.wsj.com/infectious-diseases-and-vaccines/?mc_cid=711ddeb86e, a exemplo da figura abaixo.

Measles



Note: CDC data from 2003-2012 comes from its Summary of Notifiable Diseases, which publishes yearly rather than weekly and counts confirmed cases as opposed to provisional ones.

Figura 4 - Mapa de calor

Metodologia de coleta

Os dados de desocupação podem ser obtidos de diferentes tabelas no Sistema IBGE de Recuperação Automática - SIDRA. Seguem duas tabelas que podem ser utilizadas:

Tabela 4093 - Pessoas de 14 anos ou mais de idade, total, na força de trabalho, ocupadas, desocupadas, fora da força de trabalho, e respectivas taxas e níveis, por sexo (<https://sidra.ibge.gov.br/tabela/4093>) - esta tabela exibe a taxa de desocupados já calculada e o número total da força de trabalho.

Tabela 4092 - Pessoas de 14 anos ou mais de idade, por condição em relação à força de trabalho e condição de ocupação (<https://sidra.ibge.gov.br/tabela/4092>) - esta tabela mostra os dados brutos da força de trabalho (total, ocupada e desocupada) porém sem o cálculo da taxa de desocupados.

Para fins deste trabalho foi utilizada a tabela 4092 pois contem os dados exatos em termos de quantidade e a taxa de desemprego pode ser calculada a partir dos números disponíveis. Visando a reprodutibilidade foi utilizado o pacote sidrar do R para extração dos dados. Segue código abaixo:

```
library(dplyr)
library(tidyr)
library(sidrar)

# Get data - https://sidra.ibge.gov.br/tabela/4092
ForcaTrabalho <- get_sidra(api = '/t/4092/n1/a11/n2/a11/n3/a11/v/1641/p/a11/c629/32385,32386,32387,32446')

# Add unemployed percentage
ForcaTrabalho_wide <- ForcaTrabalho %>%
  select(-'Condição em relação à força de trabalho e condição de ocupação (Código)') %>%
  spread('Condição em relação à força de trabalho e condição de ocupação', Valor) %>%
  mutate(PercentualDesocupado = `Força de trabalho - desocupada` / `Força de trabalho`)
```

```
# Export data to csv
write.csv(ForcaTrabalho_wide, "C:\\Users\\fhann\\Desktop\\\\Fernando\\Para Apple\\FGV\\
Trabalho final\\01 Dados\\PNAD\\ForcaTrabalho.csv", row.names = FALSE)
```

Os valores de desocupação foi comparada com os valores da tabela 4093 e verificou-se que são os mesmos.

Familiarização com os dados

O primeiro passo foi importar os dados para o notebook já que a visualização será realizada usando Altair no Python.

Propriedades físicas

Abaixo estão os dados iniciais, já com a coluna do percentual de desocupação. Durante a importação dos dados, a coluna "Trimestre (Código)" foi explicitamente importada como string para utilização durante a fase de exploração e modelagem.

	Nível Territorial (código)	Nível Territorial	Brasil, Grande Região e Unidade da Federação (código)	Brasil, Grande Região e Unidade da Federação	Variável (código)	Variável	Trimestre (código)	Trimestre	Unidade de Medida (código)	Unidade de Medida	Força de trabalho	Força de trabalho - desocupada	Força de trabalho - ocupada	Total	PercentualDesocupado
0	1	Brasil	1	Brasil	1641	Pessoas de 14 anos ou mais de idade	201201	1º trimestre 2012	1572	Mil pessoas	95191	7559	87632	155670	0.079409
1	1	Brasil	1	Brasil	1641	Pessoas de 14 anos ou mais de idade	201202	2º trimestre 2012	1572	Mil pessoas	96375	7245	89129	156215	0.075175
2	1	Brasil	1	Brasil	1641	Pessoas de 14 anos ou mais de idade	201203	3º trimestre 2012	1572	Mil pessoas	96454	6815	89639	156774	0.070655
3	1	Brasil	1	Brasil	1641	Pessoas de 14 anos ou mais de idade	201204	4º trimestre 2012	1572	Mil pessoas	96468	6611	89857	157426	0.068530
4	1	Brasil	1	Brasil	1641	Pessoas de 14 anos ou mais de idade	201301	1º trimestre 2013	1572	Mil pessoas	96702	7704	88999	158069	0.079667

Figura 5 - Dados

Os dados estão ordenados de forma arrumada (tidy), pronto para uso de forma geral. Algumas colunas podem ser criadas para facilitar diferentes formas de visualizações dos dados, por exemplo comparar trimestres ou anos. Também será necessário agregar dados para agrupar os estados em regiões.

Transformação dos dados

Os dados originais tiveram uma transformação inicial com a criação do percentual de desocupados conforme apresentado no código em R. Porém para possibilitar flexibilidade na visualização dos dados também é necessária transformações apresentadas a seguir.

Limpeza

Os dados originais já vieram limpos, sem necessidade de nenhuma ação neste sentido.

Conversão e Criação

Além de converter a quantidade de desocupados em percentual (taxa de desemprego), também foi necessário a decomposição da coluna "Trimestre" em número do trimestre (nome da coluna "NumTrimestre") e ano (nome da coluna "Ano").

Consolidação

A única informação não disponível nos dados originais era a qual região cada estado pertence. Apesar desta questão ser fácil de resolver manualmente designando a região correspondente a cada estado, para fins de reproducibilidade, foram extraídos os dados do site do IBGE <https://sidra.ibge.gov.br/territorio> e tratados antes de consolidar com a tabela principal dos dados. Foram agregadas as colunas "Grande Região" e "Grande Região (Código)".

Exploração dos dados

Com o objetivo de entender melhor os dados, a seguir são apresentadas as diferentes representações que servem de inspiração para a criação da visualização final. A obtenção de insights possibilita encontrar a visualização mais apropriada para estes dados.

Esta etapa de exploração também tem o objetivo de familiarizar com a nova ferramenta que estou aprendendo, o Altair do Python. Esta exploração auxilia em entender a sintaxe da linguagem, assim como os recursos disponíveis na ferramenta. O notebook utilizado para este trabalho encontra-se no seguinte link <https://colab.research.google.com/drive/1Tn0c80SpVOteGG0TKHfAVj8SprJkVJ-k?usp=sharing>.

Todas as categorias

A primeira exploração foi visualizar os dados conforme as categorias apresentadas nos dados originais: Brasil, Grande Região e Unidade da Federação. Foi utilizado o gráfico de dispersão e gráfico de linhas ao longo do tempo. É possível notar um aumento no percentual de desocupação ao longo do tempo mas também há uma flutuação da taxa por estágio.

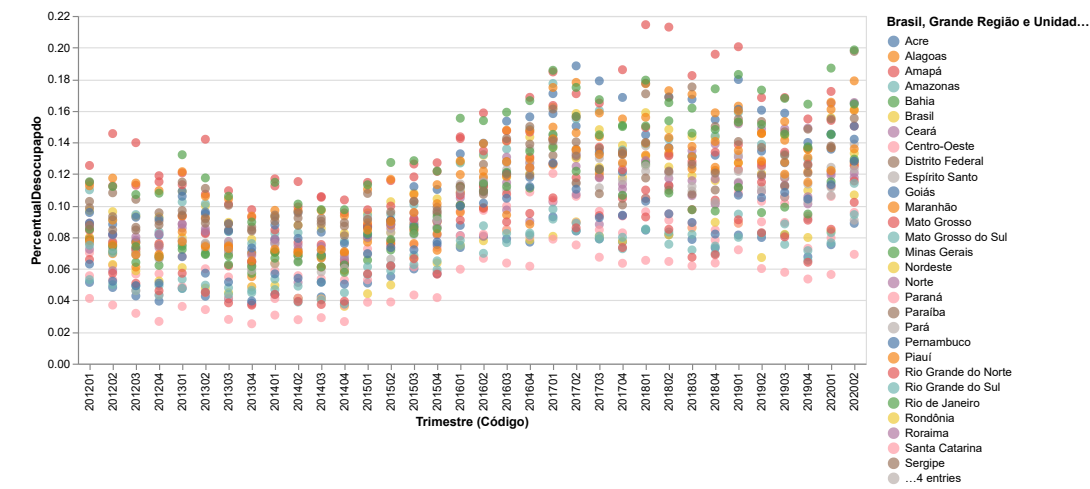


Figura 6 - Gráfico de dispersão

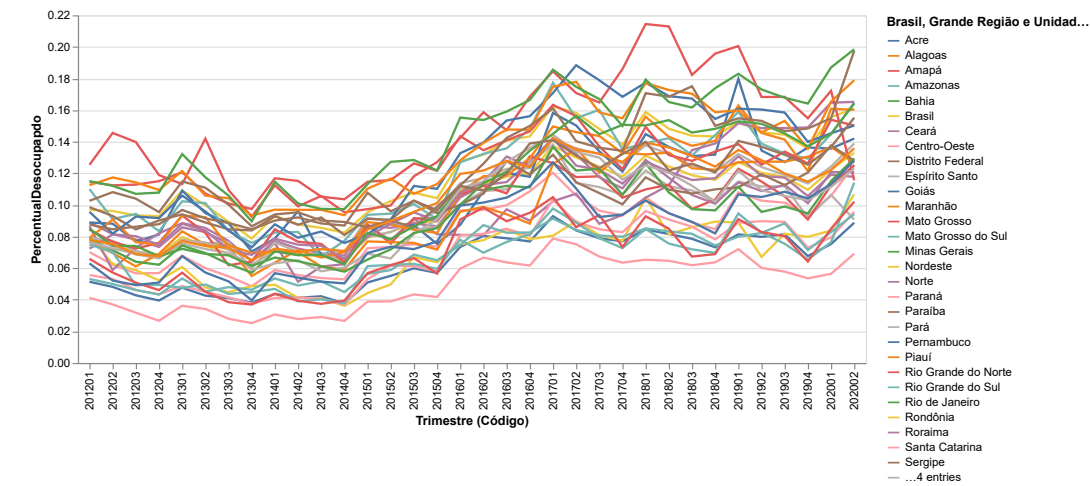


Figura 7 - Gráfico de linha

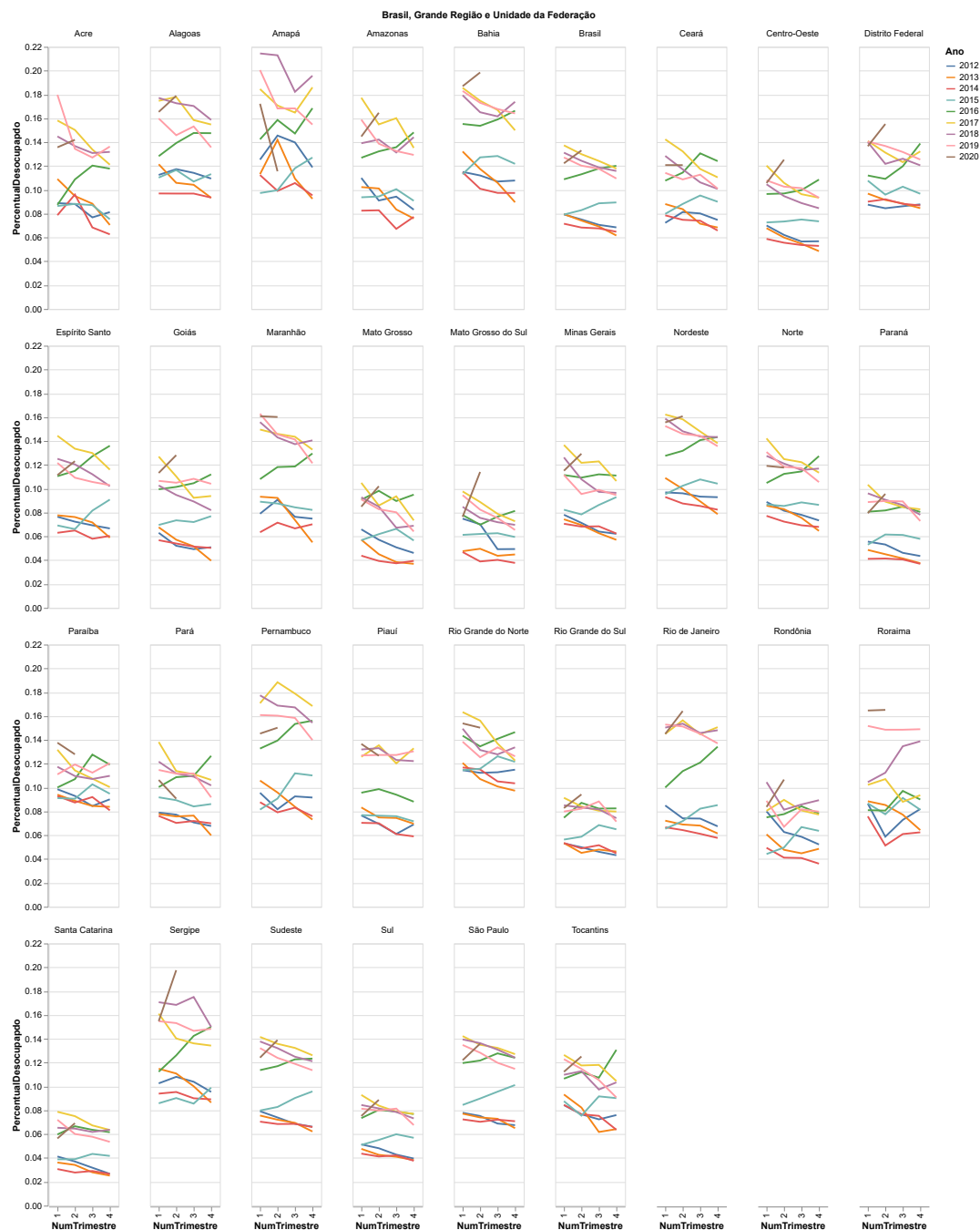


Figura 8 - Gráfico de linha em pequenos múltiplos

Agrupamento por região

A segunda exploração foi o agrupamento por região também usando gráfico de dispersão e gráfico de linha, assim como a combinação entre eles. Também foi testado o gráfico de dispersão com o tamanho dos círculos variando de acordo com a quantidade de desocupados por estado. Nesta análise é possível identificar que estados da mesma região normalmente tem taxa de desemprego próximas.

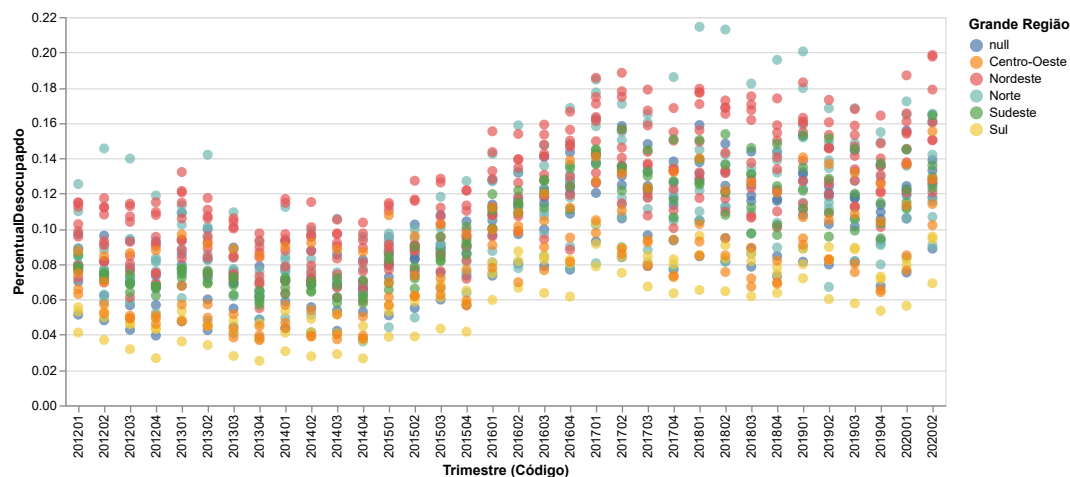


Figura 9 - Gráfico de dispersão colorido por região

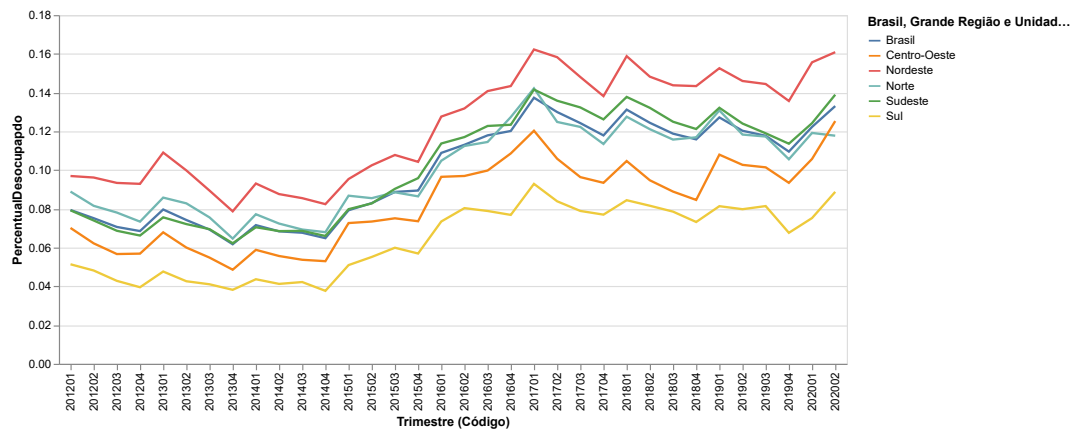


Figura 10 - Gráfico de linha por região

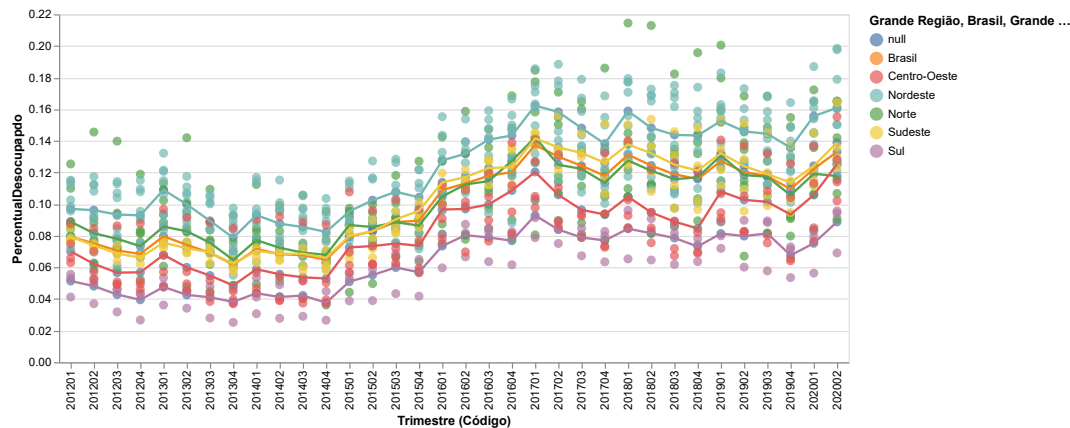


Figura 11 - Combinação gráfico dispersão e linha por região

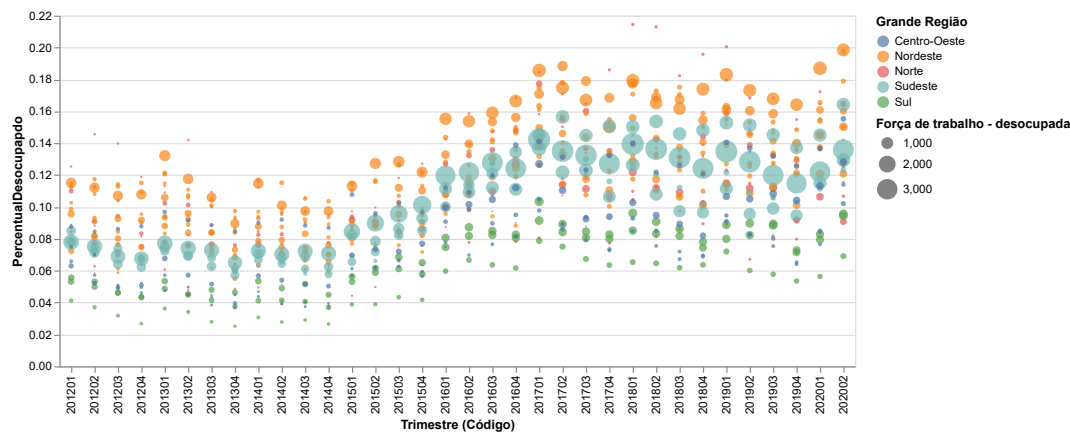


Figura 12 - Gráfico de dispersão colorido por região e tamanho proporcional a quantidade de desocupados

Interatividade da ferramenta

Finalmente, a fim de aprender mais sobre a interatividade das visualizações em Altair, foi testado o recurso de seleção para salientar uma série de dados selecionados, deixando os dados não selecionados de forma opaca, em cor cinza.

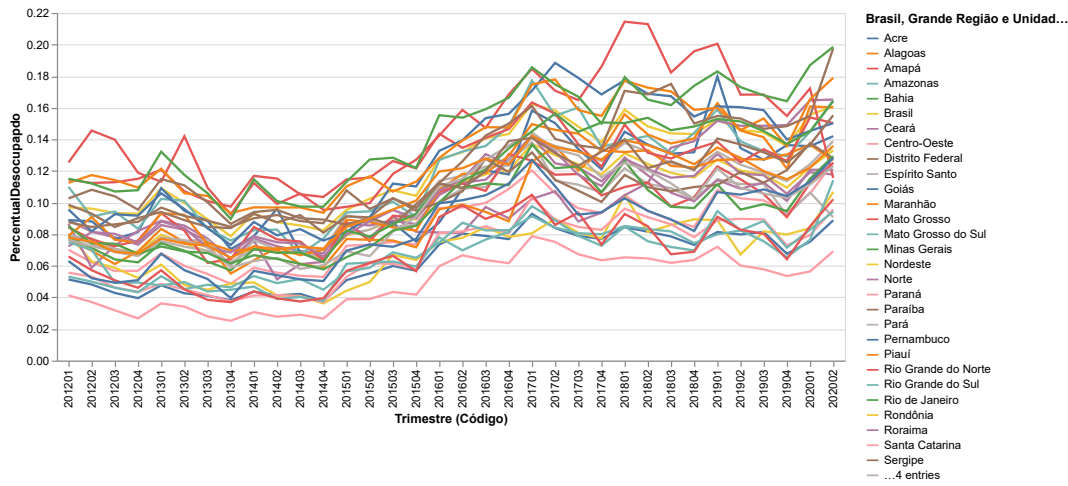


Figura 12 - Gráfico de linha apresentado anteriormente porém com interatividade acessível através deste [link](#)

Perspectiva Editorial

A fase de exploração dos dados foi importante para visualizar a taxa de desemprego de diversos ângulos e formas. Foi possível notar:

- um aumento na taxa ao longo do tempo para todos os estados,
- certa sazonalidade da taxa diminuir ao longo do ano porém com significativas exceções e,
- tendência dos estados da mesma regiões terem taxas mais próximas mas também com significativas exceções.

Baseado nestas observações a próxima etapa foi experimentar visualizar a taxa de desemprego na ideia original do trabalho, utilizando um heatmap.

Experimentação

Esta fase de experimentação foi inspirada no mapa de calor criado pelo The Wall Street Journal sobre o impacto das vacinas apresentado no início deste trabalho. Optou-se pela mesma ferramenta utilizada na fase de exploração, o Altair em Python, e foi utilizada como referência o gráfico criado pelo Franklin Oliveira, aluno do mestrado acadêmico do EMAP/FGV conforme [apresentado este link](#).

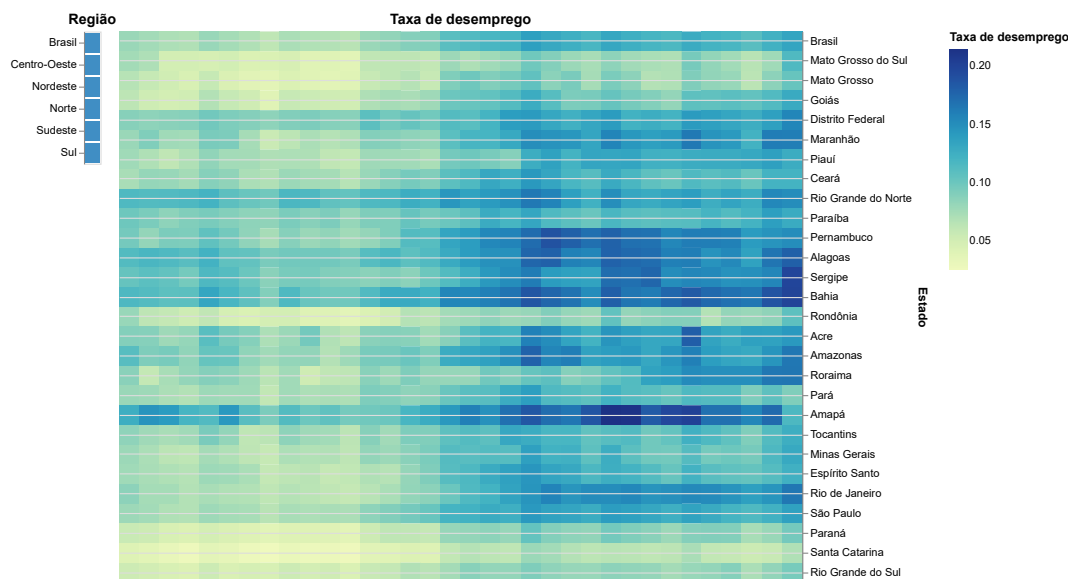


Figura 13 - Heatmap da taxa de desemprego

Este heatmap foi agrupado por região e é possível notar que para quase todos os estados a taxa de desemprego tem aumentado e que a região Nordeste tem as maiores taxas e a do Sul as menores. Durante esta análise percebe-se que não há muita informação nova comparando com os gráficos vistos na fase de exploração.

Reavaliação

Visto que a visualização da taxa de desemprego não gerou muitos insights, foram repensados dois aspectos que poderiam tornar a visualização mais coerente e atrativa para análise.

1. **Variação trimestral da taxa de desemprego:** dado que o aumento da taxa de desemprego ao longo dos anos não é novidade, um dos fatores que possibilita o entendimento da situação do desemprego a cada momento é visualizar a diferença da taxa de desemprego de um trimestre para o outro. Desta forma é possível perceber um aumento ou declínio na taxa conforme apresentado na figura abaixo. Apesar desta alteração, ainda não é possível visualizar tendências relevantes.

Desemprego em estados acompanham crises de diferentes formas e intensidade

Crise econômica e política desencadeada em meados 2014 e início da pandemia pela Covid-19, veja as correlações. Roraima aparenta ser outro país. (Taxa = percentual de desocupados, Variação = taxa do trimestre atual - taxa do trimestre anterior * 100, Fonte = PNAD Contínua, IBGE)

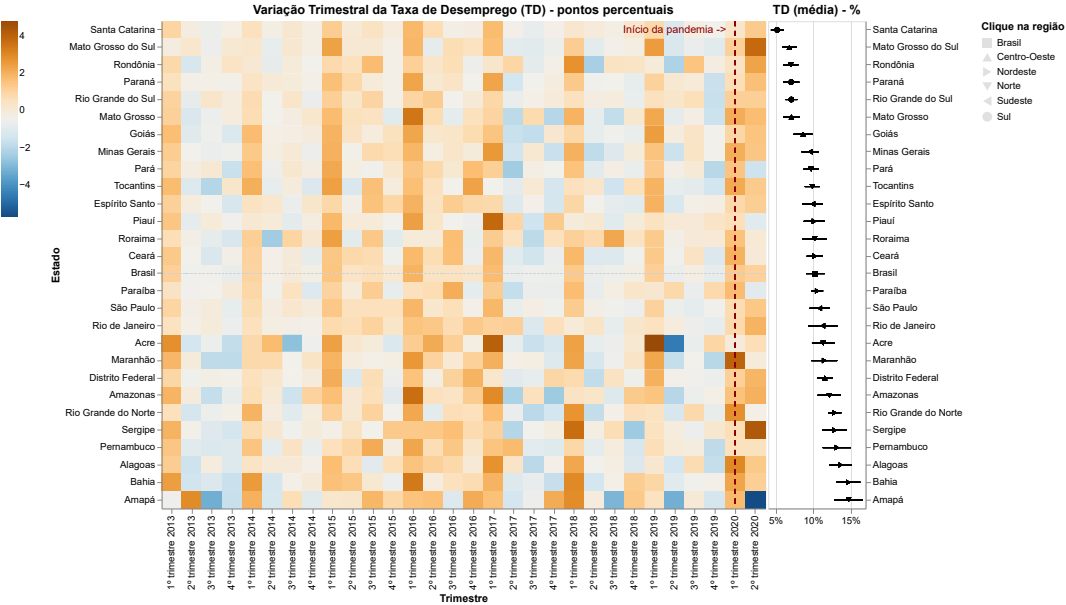


Figura 14 - Diferença entre trimestres da taxa de desemprego

2. Média móvel da taxa de desemprego: a solução para permitir visualização de tendências utilizando a diferença da taxa de desemprego entre trimestres foi calcular a média móvel, que suaviza flutuações de curto prazo, leva em consideração efeitos de sazonalidade e destaca tendências de longo prazo. Foi utilizada uma média móvel simples baseada no período de um ano, isto é, 4 trimestres. Trata-se de um indicador mais "lento" mas que pode ser mais robusto para análise exploratória.

Força de Trabalho Desocupada no Brasil

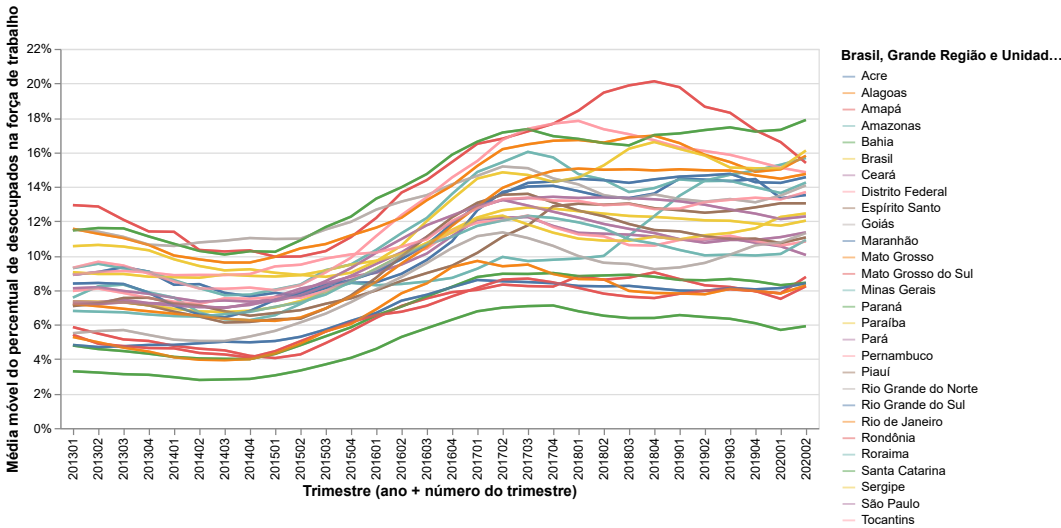


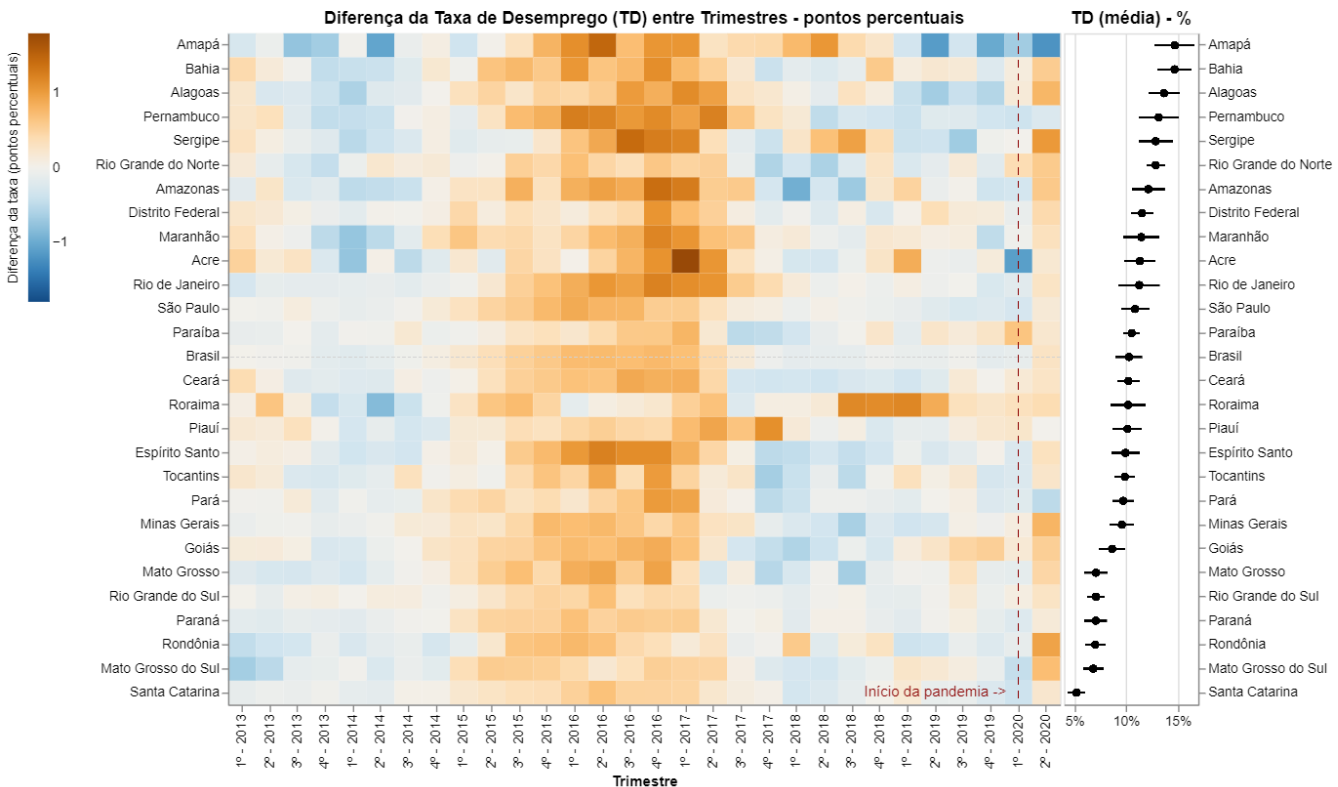
Figura 15 - Média móvel da taxa de desemprego

Visualização Final

Com as reavaliações implementadas, a **visualização final** é apresentada a seguir, incluindo perspectiva editorial para o leitor alvo deste trabalho.

Desemprego em estados acompanham crises de diferentes formas e intensidade

Crise desencadeada em meados 2014 e início da pandemia pela Covid-19 em 2020, veja as correlações. Amapá tem maior média da taxa de desemprego e Roraima aparenta ser outro país. (Taxa = média móvel do percentual de desocupados de 4 trimestres, Diferença = taxa do trimestre atual - taxa do trimestre anterior * 100, Fonte = PNAD Contínua, IBGE)



Escolhas Gráficas

O tipo de gráfico selecionado foi um mapa de calor, heatmap, que é capaz de representar várias categorias, estados brasileiros, para um mesmo dado numérico, taxa de desemprego. Este tipo de representação permite seguir o mantra de visualização da informação segundo Shneiderman:

1. **Overview:** uma visão geral da situação em todo o país em diversos momentos no tempo,
2. **Zoom:** é possível focar em um estado assim como período para todos os estados,
3. **Filter:** o menu de seleção acima do gráfico permite ver os estados por regiões,
4. **Details on-demand:** e finalmente o recurso de tooltip mostra os detalhes para cada ponto no gráfico.

Foi incluso um gráfico auxiliar à direita para mostrar a média da taxa de desemprego por estado, a qual foi utilizada para ordenar os estados, de cima para baixo, do maior para o menor. Este gráfico visa auxiliar em visualizar a faixa de taxa de desemprego durante os anos, complementando assim a informação do mapa de calor.

Representação dos Dados

No mapa de calor os dados são representados por cores e saturação das mesmas. Foi escolhida uma paleta de cores divergentes para facilitar a interpretação de quando a taxa de desemprego aumenta (laranja) e diminui (azul), sendo que a saturação indica a intensidade em ambos os sentidos. Neste caso o objetivo é mostrar comportamento de forma geral, caso o leitor queira saber mais detalhes pode utilizar o tooltip para visualizar os dados exatos de cada ponto.

No gráfico auxiliar os dados foram representados por pontos com posicionamento em relação ao eixo x. A barra de erro em cada ponto representa o desvio padrão da taxa de desemprego que visualmente mostra ao leitor a faixa significativa dos valores para cada estado.

Interatividade

O uso do Altair como ferramenta para criação da visualização foi fundamental para permitir a interatividade. O recurso do tooltip permite visualizar os dados pontualmente quando o leitor quiser mais detalhes. O filtro através de seleção permite destacar os grupos desejados, neste caso forma os estados de cada região. Este filtro é aplicado ao gráfico principal assim como o auxiliar para focar a análise em grupos separadamente em busca de tendências, padrões e outliers. Os estados que não estão na região selecionada são transformados na cor cinza.

Anotações

Para o eixo y foram utilizados os nomes completos dos estados a fim de clareza já que este tipo de visualização permite textos mais longos. A escolha do eixo vertical aparecer nos dois gráficos, principal e auxiliar, foi para auxiliar na leitura já que cada linha do mapa de calor é extensa. Quanto ao texto do eixo x foi reduzido ao máximo, sem comprometer o entendimento, já que a orientação do texto está na vertical, portanto mais difícil a leitura de textos longos.

No título do gráfico principal foi colocado TD entre parênteses para abreviar taxa de desemprego para que esta abreviação fosse usada no título do gráfico auxiliar já que a largura é menor. Para o título do selecionador foi escolhido um comando, "Selecione região", para o leitor perceber que é possível interagir com esse filtro. Infelizmente a ferramenta coloca automaticamente um traço inferior (underline) entre uma palavra e outra e também não aceita acentos.

O título foi elaborado para atrair o leitor para o gráfico, já enfatizando o dado que o gráfico representa e a correlação com as crises, para direcionar a leitura. No subtítulo foi dado mais detalhes e uma curiosidade, sobre Roraima, para atrair o leitor a engajar com o gráfico e facilitar o entendimento no primeiro momento. No segundo subtítulo foram explicadas as variáveis e devido a limitação da ferramenta, também foi adicionado a fonte dos dados que normalmente é no rodapé do gráfico.

Finalmente, no mapa de calor também foram incluídas duas marcações para servir de referência na leitura do gráfico. A marcação horizontal foi uma linha tracejada bem leve para média do Brasil que serve como referência para ver os estados que estão acima ou abaixo da média nacional. A marcação vertical encontra-se no 1º trimestre de 2020 que foi quando foi decretada a pandemia pela Organização Mundial da Saúde (OMS) da ONU em 11 de março de 2020. Esta marcação foi colocada em cor vermelha, assim como o texto indicando que a linha refere-se ao início da pandemia.

Cores

A escolha de uma paleta divergente foi para evidenciar uma mudança, aumento e diminuição como descrito anteriormente. Foram adotadas cores contrastantes para enfatizar esta divergência, onde normalmente o azul é associado a frio e menos e o laranja mais a quente e alto. Estas cores também são inclusivas para aqueles que tem habilidade reduzida para distinguir entre cores.

As cores do título e subtítulo seguiram as cores da paleta do mapa de calor a fim de evidenciar uma associação entre o gráfico e o título. O gráfico auxiliar foi mantido na cor preta para não distrair o leitor da visualização principal.

Composição

Quanto a composição da visualização optou-se por um título e subtítulo geral, que se refere tanto ao gráfico principal quanto ao auxiliar, na parte superior.

O gráfico principal, o mapa de calor, foi posicionado do lado esquerdo já que no ocidente a leitura é da esquerda para a direita portanto dando ênfase a este gráfico no primeiro momento. A legenda inicialmente foi fixada do lado direito, porém com a inclusão do gráfico auxiliar no lado direito, a legenda foi transferida para a esquerda.

Finalmente o menu selecionador das regiões foi disposto entre o subtítulo e o gráfico principal para ser facilmente visível no início da leitura da visualização.

Validação

Durante o curso a professora e os colegas apoiaram na análise da visualização com comentários e sugestões que permitiram a evolução contínua do trabalho. Com a base teórica ao longo do curso foi possível analisar e ver soluções de outras pessoas que auxiliaram na análise crítica e criatividade.

Para o processo de validação optou-se pela análise qualitativa de pessoas com o perfil do público alvo desta visualização e experts em áreas correlatas. O número adequada do feedback para visualizações praticado normalmente tem sido sete. Esta visualização portanto foi submetida para sete pessoas com as seguintes qualificações:

- 3 economistas,
- 2 engenheiros,
- 1 contadora,
- 1 designer.

De modo geral os economistas foram cruciais analisar as taxas utilizadas e a forma de cálculo da média para entender potenciais aplicações. Também deram feedback sobre interpretação e usabilidade que foram importantes para adequar as anotações. Os engenheiros e contadora comentaram mais sobre a experiência do usuário e potencialidades do uso da visualização. O designer foi importante entender melhores formas de formatar e compor a visualização de forma mais intuitiva.

As principais descobertas no processo de validação foram:

1. Indicador utilizado é um indicador "lento", que mostra tendência a longo prazo. O indicador utilizado foi a média móvel simples. Para fins deste trabalho foi mantido este indicador.
2. Inicialmente foi usado termo "variação trimestral" no título do mapa de calor, o que não causou muita dúvida, principalmente dos economistas. Porém para ficar ainda mais claro o que o dado se refere, este termo foi substituído por "diferença entre trimestres" para tornar mais direta e rápida a compreensão.
3. Uma das observações foi em relação à legenda que o máximo do valor negativo não era o mesmo do máximo do valor positivo pois foi utilizada a configuração padrão da ferramenta. A alteração para colocar o valor máximo negativo e positivo o mesmo não foi só para uniformizar a legenda mas também tornou zero o ponto de divergência das cores, algo que era esperado mas foi somente detectado por acaso por uma pessoa.
4. Apesar da existência do gráfico auxiliar, muitos tiveram dificuldade em notar rapidamente que o ordenamento era conforme a média da taxa de desemprego por estado. Uma primeira alteração foi inverter o ordenamento para decrescente pois desta forma fica coerente se referir aos estados que estão posicionados acima da média nacional estarem acima do item Brasil. Além disso, foi adicionado no subtítulo que Amapá é o estado que tem a maior média da taxa de desemprego no país, chamando a atenção para o motivo de aparecer primeiro e portanto deduzir que os demais estão na ordem decrescente.
5. Praticamente todos não identificaram o menu seletor mesmo depois que o título foi alterado para um comando "Clique na região". Este menu ficava no lado direito e era uma lista com o nome das regiões porém não havia um indicativo claro de interação. A solução foi alterar para um seletor em formato de botões de rádio, radio buttons, que os leitores estão mais acostumados. Um desafio, porém, foi que o Altair adiciona este seletor abaixo do gráfico que não fica acessível, o leitor provavelmente nem notaria. Não existe comando na ferramenta para alterar isso portanto a solução foi alterar o HTML para reposicioná-lo para o topo do gráfico inserindo o seguinte código:

```
# Código HTML para reposicionar seletor colocado dentro do tag: style

form.vega-bindings {
  position: absolute;
  right: 630px;
  top: 63px;
  font-size: 70%;
  font-family: Arial;
}
```

Conclusão

Este trabalho foi fundamental para aplicação teórica dos conceitos apresentados em sala de aula com um objetivo pessoal que fez com que a experiência fosse ainda mais enriquecedora. A troca de feedback dentro e fora da sala de aula ajudou a ver na prática a eficácia das escolhas gráficas. Passar por todo o processo de criação e análise permitiu oportunidades de aprendizados com erros e acertos.

Esta visualização teve um objetivo mais geral e exploratório. Em futuros trabalhos o desafio será adequar a visualização para objetivos específicos dependendo da linha editorial escolhida. É possível a aplicação por exemplo de indicadores mais ágeis para serem utilizados para previsão e tomada de decisão, por exemplo através do uso de média móvel exponencial ao invés da simples aplicada neste trabalho. Outro desafio é melhorar a experiência do leitor da visualização através de objetos gráficos mais intuitivos que permitam ao leitor explorar o máximo da visualização.

Por fim, gostaria de agradecer a professora Asla por todo o compartilhamento do conhecimento e disponibilidade, assim como todos os colegas do curso. Também agradeço às sete pessoas que me auxiliaram na avaliação da visualização e a minha esposa e filha por me apoiarem nesta disciplina.

Caso queiram entrar em contato comigo, disponibilizo aqui [meu email](#) e [perfil do LinkedIn](#). Muito obrigado e espero que este trabalho sirva de inspiração e aprendizado.

-
- Esta página está disponível no site <https://fehann.github.io/VisEMAp2020Trabalho/>
 - Para ver somente a visualização final clique aqui <https://fehann.github.io/VisEMAp2020Trabalho/html/chartfinal1.html>
 - Todos os arquivos, incluindo notebook, estão disponíveis neste site <https://github.com/fehann/VisEMAp2020Trabalho>