

Trabalho Final

Professor(a): Asla Medeiros e Sá

Aluno: Franklin Alves de Oliveira

Casamentos e divórcios: uma análise visual para países da OCDE

Resumo

Trabalho final da disciplina de Visualização de Dados do curso de Mestrado em Modelagem Matemática da Informação da EMap/FGV.

Esse trabalho tem como objetivo aplicar técnicas e ferramentas de visualização apresentadas ao longo do curso. Pessoalmente, escolhi aplicar técnicas de visualização de dados a um *dataset* de taxas de casamento e divórcio de países integrantes da OECD¹. A ferramenta escolhida para a execução deste trabalho foi a biblioteca Altair² para Python 3.

O relatório está dividido da seguinte forma: Na seção 1 são apresentadas algumas motivações que levaram à escolha desse tema; a seção 2 apresenta os datasets que foram criados a partir dos dados coletados da OECD e que deram suporte à criação dos gráficos; a seção 3 exibe um *screenshot* das visualizações criadas e alguns comentários relacionados à identificação de possíveis padrões; esse relatório é finalizado apresentando as referências que influenciaram fortemente o processo criativo.

1 Introdução

Esse projeto de visualização tem a intenção de aplicar técnicas e ferramentas voltadas para visualização de informações (incluindo dados) apresentadas e discutidas ao longo do curso de Visualização de Dados para o mestrado acadêmico da Escola de Matemática Aplicada da FGV³.

Para a realização desse projeto, foi escolhido um *dataset* de taxas de casamento e divórcio para países da OECD. O arquivo original pode ser encontrado no link apresentado no rodapé desta página. Nesse *dataset*, são encontradas taxas de casamento e divórcio⁴ e também a idade média de homens e mulheres em seu primeiro casamento, para um total de 45 países.

A proposta definida para esse *dataset* foi de explorar a existência de uma possível tendência temporal no que diz respeito à forma casamento tradicional. Com o passar dos anos, novas formas contratuais de relacionamento foram surgindo como por exemplo, a união estável. Com isso, buscamos investigar se há uma diminuição no número de casamentos tradicionais ao longo do tempo.

A seguir, é exibido um gráfico para da taxa de casamentos nos Estados Unidos a partir de 1990 (o link para a versão publicada em html está disponível no arquivo [usa_marriage_rate](#)), originalmente apresentado e disponível nesse link⁵, também presente no rodapé desta página. Nele, é fácil perceber a tendência de redução no número de casamentos tradicionais a partir dos anos 90. Com isso, naturalmente surge a pergunta: "Será que esse padrão também é perceptível para outros países?"

¹Link para download do dataset: <http://www.oecd.org/els/family/database.htm>. Para esse trabalho, foi utilizado o conjunto de dados denominado "SF3.1 Marriage and divorce rate".

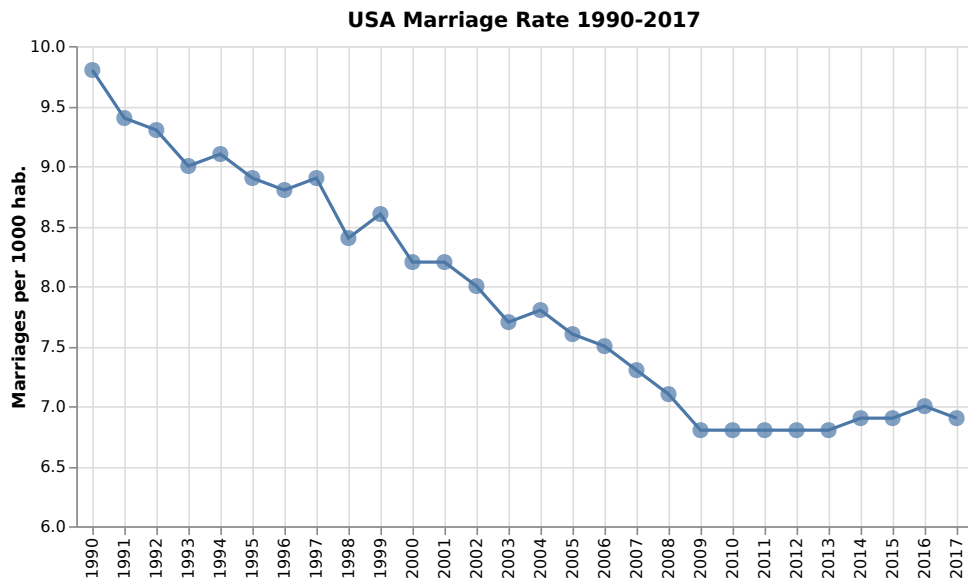
²Altair docs: <https://altair-viz.github.io/>

³Fundação Getúlio Vargas - Rio de Janeiro.

⁴Taxa de casamentos: N° de casamentos/1000 hab.; Taxa de divórcios: N° de divórcios/1000 hab.

⁵<https://www.statista.com/statistics/195951/marriage-rate-in-the-united-states-since-1990/>

Figura 1: Taxa de casamentos nos EUA (1990 a 2017)

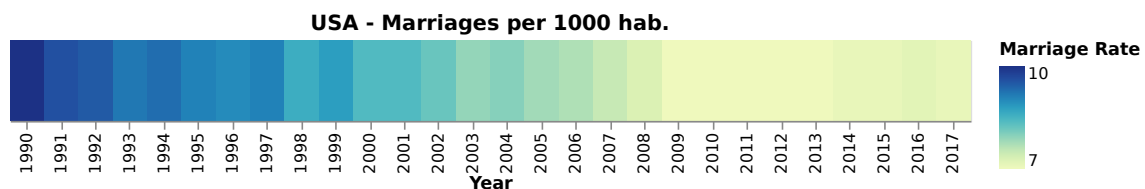


A partir dessa informação, outras perguntas vão surgindo de maneira natural. Como, por exemplo, "como as taxas de casamento e divórcio estão associadas? Será que um menor número de casamentos ocasiona em uma menor taxa de divórcios? Ou será que há uma redução no número de casamentos devido ao número de divórcios?"⁶

São perguntas como essas que buscaremos responder nesse estudo.

OBS: por fins de ilustração da representação gráfica que usaremos nas próximas visualizações, vamos representar a mesma informação, codificando a taxa de casamentos (eixo Z) no campo da cor. O resultado pode ser conferido abaixo e aqui está o link para a representação gráfica em html.

Figura 2: Taxa de casamentos nos EUA (1990 a 2017) na representação gráfica escolhida



Com essa representação, esperamos realçar tendências de elevação/diminuição para um grande conjunto de dados⁷.

2 Dataset

Nesta seção, apresentaremos brevemente os *datasets* utilizados nesse trabalho. A partir dos dados coletados da OECD, foram construídas as seguintes tabelas:

Na figura 3, temos as taxas de casamento (Número de Casamentos/1000 hab.) identificadas por país, ano e continente. Na figura 4, temos as taxas de divórcio (Número de Divórcios/1000 hab.) também identificadas por país, ano e continente.

⁶Por haver uma maior ruptura nesses contratos matrimoniais, as novas gerações podem estar aderindo mais a outras formas de relacionamento de longo prazo.

⁷Apesar de a cor não retratar a quantidade de maneira tão precisa quanto no primeiro gráfico, é eficiente em realçar tendências e destacar padrões ao visualizar um grande volume de dados em um só gráfico. É por esse motivo que escolhemos essa representação.

Figura 3: Taxas de casamentos

	Country	Year	marriage_rate	Continent
1904	China	2010	9.3	Asia
1905	China	2011	9.3	Asia
1906	China	2012	9.6	Asia
2034	Cyprus	1964	4.4	Asia
2035	Cyprus	1965	7.6	Asia

Figura 4: Taxas de divórcios

	Country	Year	divorce_rate	Continent
1706	China	2010	2.0	Asia
1707	China	2011	1.6	Asia
1708	China	2012	1.8	Asia
1828	Cyprus	1964	0.2	Asia
1829	Cyprus	1965	0.2	Asia

Por fim, na figura 5, temos as idades médias em que Homens e Mulheres casaram pela primeira vez, identificadas por país, ano e continente. A partir dessas idades, foi construída a coluna de idade média do casal no primeiro casamento.

Figura 5: Idades médias no primeiro casamento

	Country	Year	male_mean_age	female_mean_age	couple_mean_age	Continent
890	Cyprus	1990	27.9	24.9	26.40	Asia
891	Cyprus	1991	27.5	24.6	26.05	Asia
892	Cyprus	1992	27.9	25.4	26.65	Asia
893	Cyprus	1993	27.9	25.2	26.55	Asia
894	Cyprus	1994	28.4	25.8	27.10	Asia

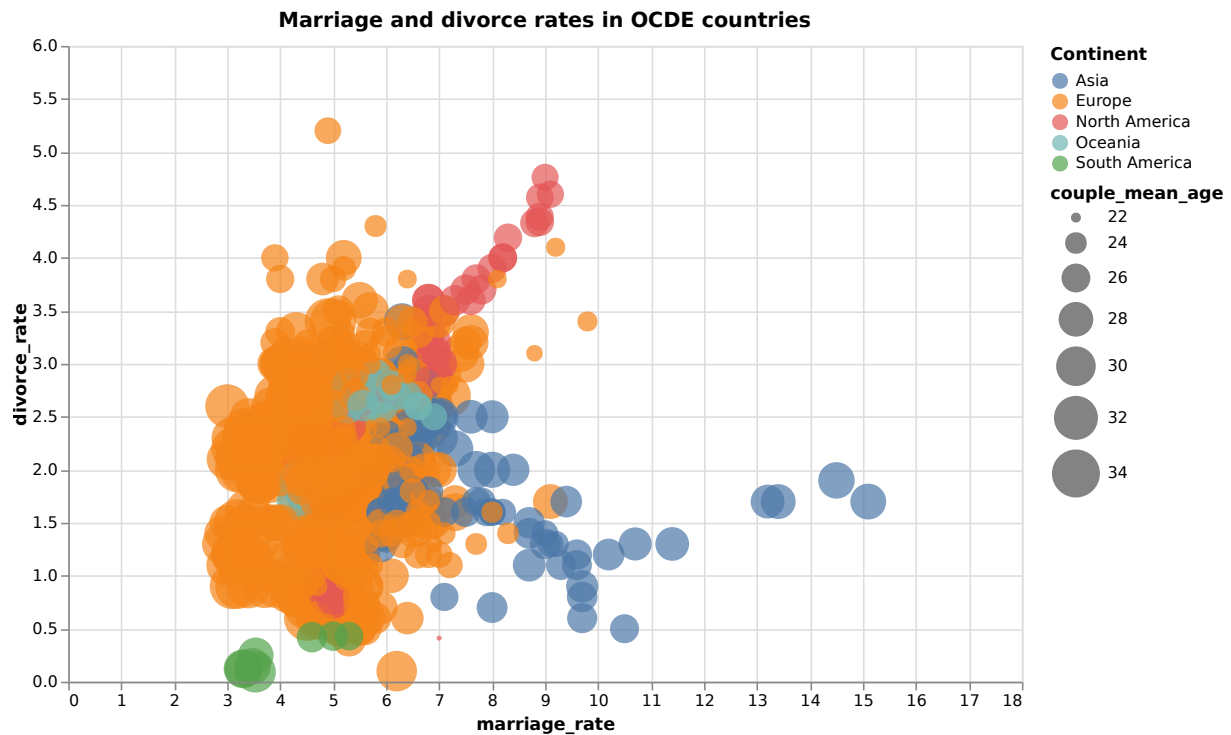
3 Visualizações

Nessa seção, serão apresentadas as propostas de visualização para os *datasets* apresentados na seção 2. Com esses gráficos, esperamos responder algumas perguntas a respeito de possíveis tendências temporais envolvendo a forma de casamento tradicional para 45 países membros da OECD.

3.1 Proposta gapminder-like

A primeira proposta de visualização foi fortemente inspirada na famosa apresentação feita por Hans Rosling [4] em seu famoso TED Talk apresentado em 2007. De maneira semelhante, procurou-se explorar a tendência temporal entre casamentos (eixo X) e divórcios (eixo Y). A seguir, é apresentada uma figura estática desse gráfico.

Figura 6: Gráfico gapminder-like para taxas de casamento e divórcio



Nesse gráfico estilo gapminder, o tamanho do círculo foi codificada como a idade média em que os casais casaram-se pela primeira vez, e a cor como o continente. A versão dinâmica deste gráfico é apresentada nesse link⁸: [Gráfico gapminder-like dinâmico](#).

⁸arquivo gapminder_like.html

Apesar de os pontos se moverem ao longo do tempo para um grupo de maior concentração, não foi possível extrair um padrão visual muito claro das tendências temporais com essa proposta de visualização (pelo menos não um padrão tão perceptível quanto no gráfico original).

Vale ressaltar que, apesar de o padrão não ser tão perceptível dessa forma, outras construções gráficas podem torná-los mais perceptíveis. Com esse pensamento, foram formuladas as representações gráficas apresentadas a seguir para cada uma das variáveis individualmente. Vale destacar que essas visualizações foram formuladas com base no processo de construção de um projeto de visualização apresentado por Andy Kirk em seu livro "Data visualization" [3] e nos elementos de construção de gráficos genialmente apresentados por Jacques Bertin em seu livro "Semiology of graphics" [1].

3.2 Gráfico 1: Something borrowed, something...

Nesse gráfico, buscou-se analisar as taxas de casamento para os 45 países presentes no conjunto de dados de 1960 a 2017. Ao codificar o tempo no eixo X, os países no eixo Y (agrupados e ordenados por continente) e as taxas de casamento no campo da cor, em uma escala de cores em azul⁹, foi possível detectar uma tendência de redução no número de casamentos para esses países. Esse fator pode estar associado à adoção de novas formas de relacionamento estável, que não o casamento tradicional.

Vale destacar que esse gráfico herda elementos das famosas visualizações apresentadas pelo *Wall Street Journal*¹⁰ [2] para, visualmente, comprovar a eficácia de vacinas para várias doenças. O link para essa visualização é disponibilizado no rodapé.

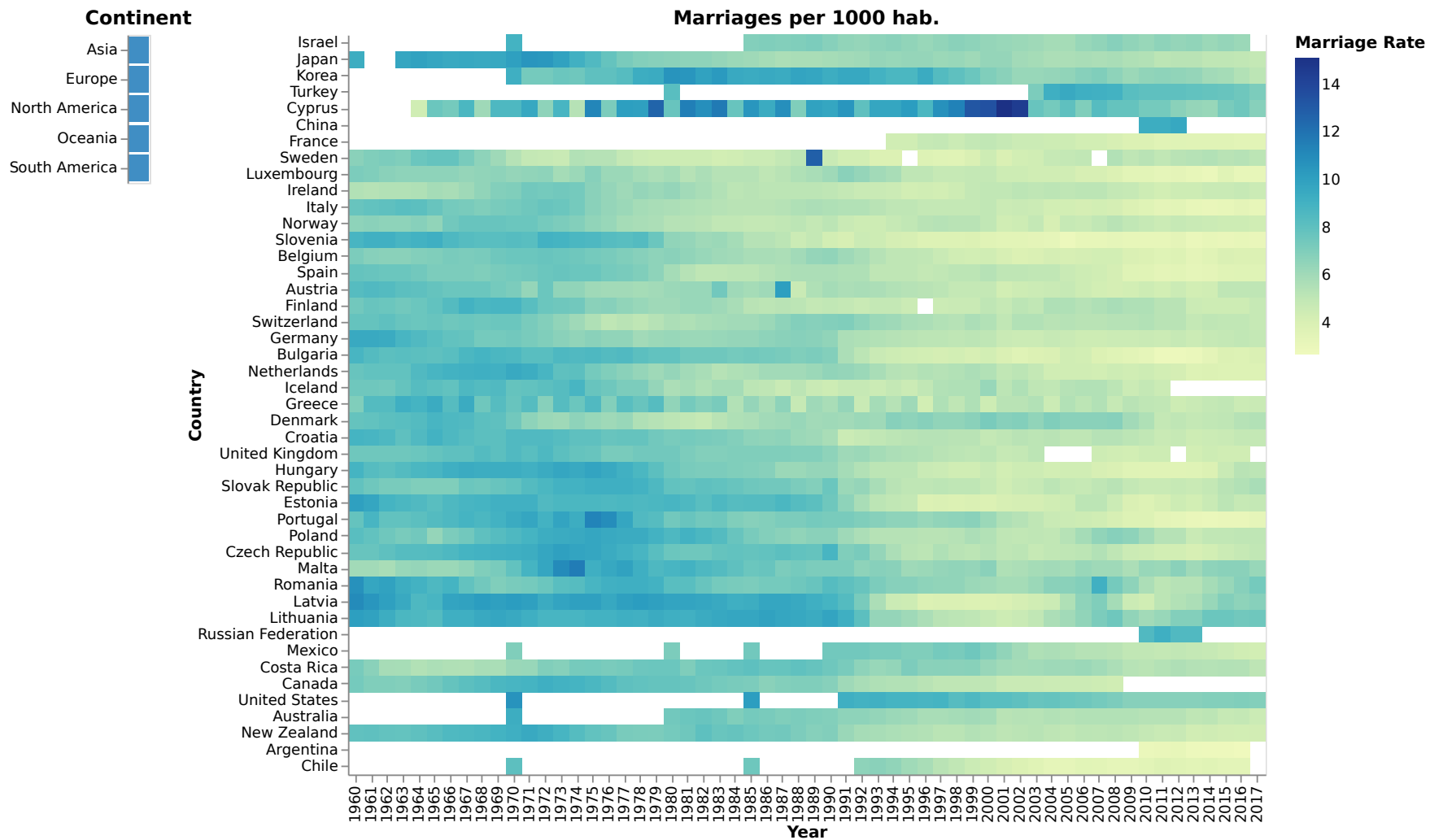
Para melhor visualizar esse padrão por continentes (possivelmente identificar padrões culturais), foi adicionada uma camada de interatividade com filtros de seleção. A versão interativa desse gráfico está disponível nesse link¹¹: [Something borrowed, something...](#)

⁹para fazer referência à tradição norte americana "something old, something new, something borrowed and something blue".

¹⁰Link:<https://flowingdata.com/2015/02/19/impact-of-vaccines-throughout-history/>

¹¹Arquivo marriages.html

Figura 7: Taxas de casamento para países da OECD (1960-2017)



3.3 Gráfico 2: Now and for... ever?

De maneira análoga ao apresentado para as taxas de casamento, na próxima página é exposto um gráfico semelhante, codificando a cor como as taxas de divórcio, em uma escala de cores marrom¹². Ao contrário do que foi constatado para a variável do casamento, as taxas de divórcio estão visualmente elevando-se ao longo do tempo, mas a uma velocidade bastante distinta para alguns países.

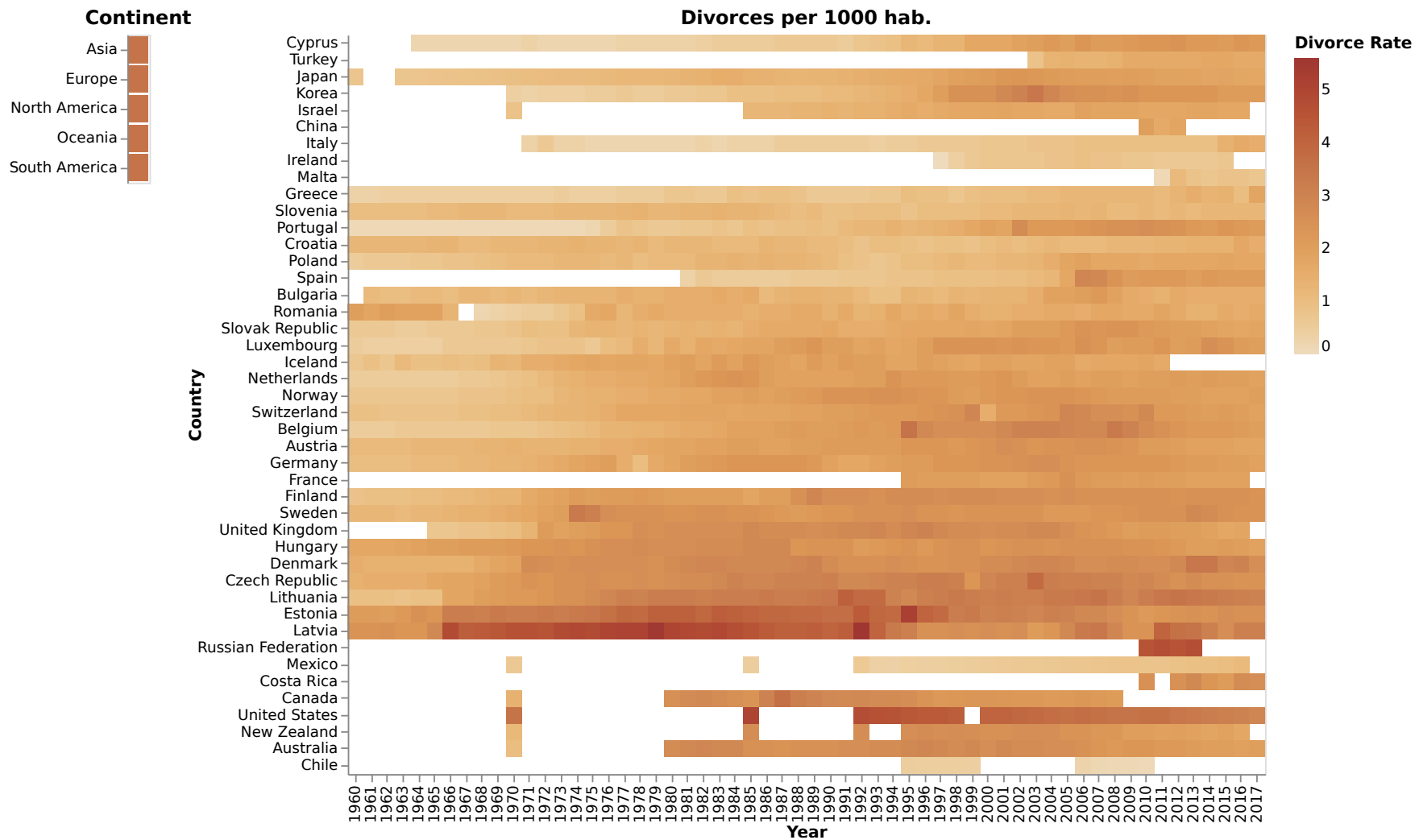
Uma possível explicação para isso são fatores culturais associados ao divórcio. Algumas culturas vêem essa prática negativamente. Outro fator pode ser o fato de que alguns países passaram a permitir legalmente essa prática tardiamente, e o crescimento visualizado na imagem é efeito das sociedades, aos poucos, reconhecendo e aderindo a essa prática.

Aqui é apresentada a versão interativa desta visualização¹³: [Now and for... ever?](#).

¹²Essa escala de cores foi escolhida para representar os dados de maneira neutra. A ideia não é apresentar o divórcio como algo bom ou ruim, mas sim um retrato imparcial do que aconteceu nesses 45 países.

¹³[arquivo divorces.html](#)

Figura 8: Taxas de divórcio para países da OECD (1960-2017)



3.4 Gráfico 3: Em média, as pessoas estão casando mais tarde.

Por fim, vamos explorar a informação da idade média na qual as pessoas (homens e mulheres) casaram-se pela primeira vez. Essa informação é codificada no eixo X. No eixo Y, mais uma vez, temos os países agrupados por continente. O gênero (masculino e feminino) estão representados em escalas de cores distintas: azul e rosa, respectivamente. O tempo, em anos, é representado na intensidade das cores¹⁴ (cores mais escuras representam períodos mais recentes). Esse gráfico é exibido na página seguinte e a sua versão interativa está disponível por meio desse link:¹⁵ [Idade média no primeiro casamento](#).

Analisando o padrão visual calmamente, é possível facilmente detectar um padrão: a idade média na qual homens e mulheres estão casando pela primeira vez está se elevando com o passar dos anos (isto é, pontos mais escuros estão localizados mais à direita). Incrivelmente, esse é um padrão que persiste a todos os países da amostra.

Para melhor visualizar essa tendência, foi adicionada à esse gráfico uma animação à la Hans Rosling (i.e., um slider para visualizarmos um ano por vez.). O resultado dessa animação é apresentado aqui¹⁶ [Idade média no primeiro casamento, animado](#). É facilmente notável que, com o passar do tempo, os pontos movem-se para a direita, tornando a proposição de que as pessoas, em média, estão casando mais tarde irrefutável¹⁷.

Além disso, podemos nos atentar para o fato de que, apesar de homens e mulheres estarem postergando o ato de casar-se, a diferença entre as idades dos casais, em média, tende a permanecer a mesma com o passar dos anos. No gráfico, esse padrão é percebido como os pontos na cor azul e rosa estarem dando passos de tamanhos semelhantes para a direita, de forma que estes aparentam estar equidistantes com o passar do tempo.

Por fim, outro padrão interessante é que, em países com tradição religiosa mais forte (como Turquia e Israel), a idade média em que homens e mulheres se casaram pouco mudou com o passar dos anos.

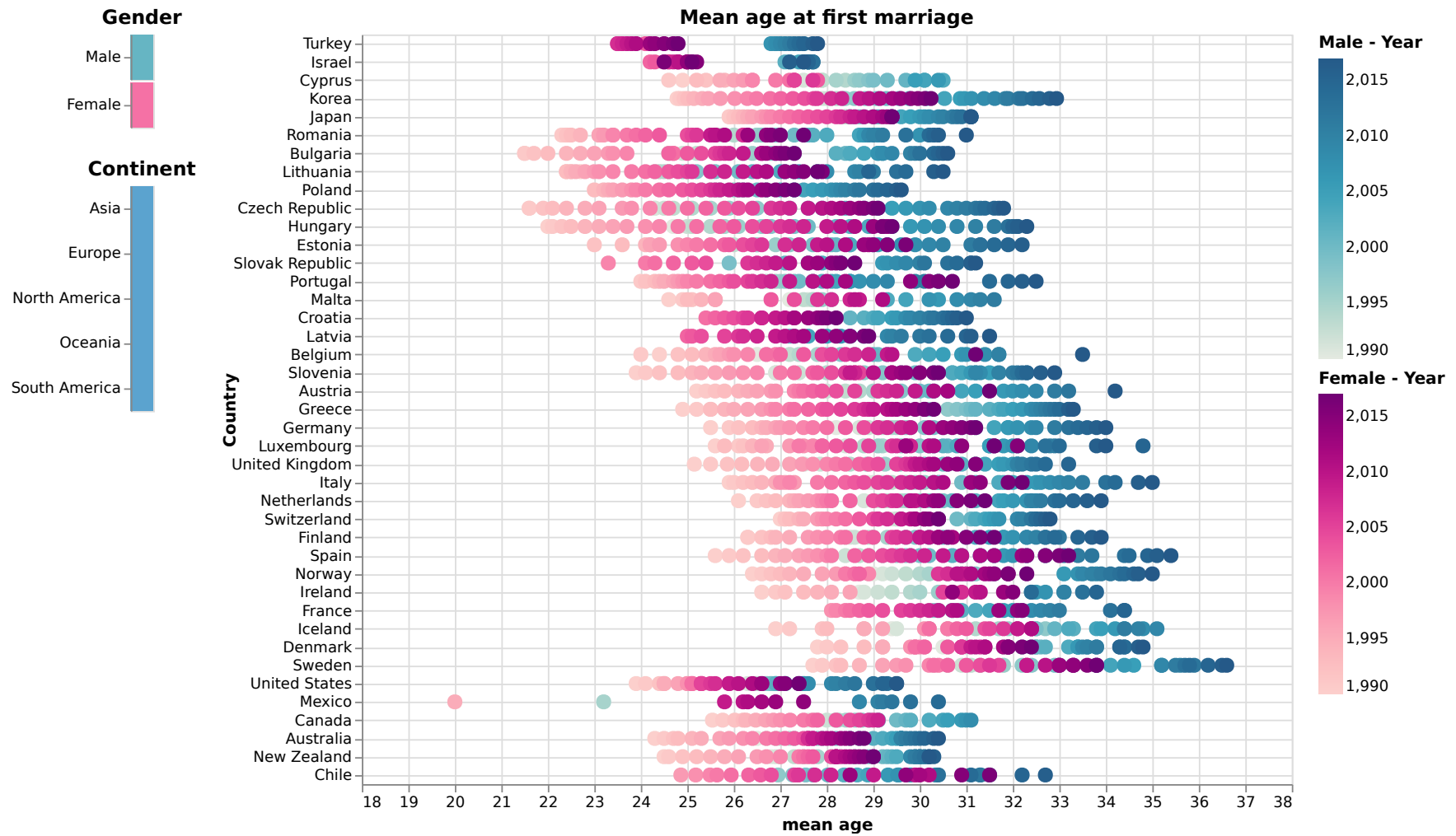
¹⁴Idealmente, essa representação deve ser feita em uma escala de cores discreta, dado que o tempo é uma variável discreta. No entanto, para o gráfico apresentado aqui, foram adotadas escalas de cor contínuas, uma vez que as escalas discretas excediam o limite da tela delimitado para a visualização.

¹⁵[arquivo mean_age.html](#)

¹⁶[arquivo mean_age_animated.html](#)

¹⁷para os 45 países que compõem o dataset da OECD.

Figura 9: Idade média no primeiro casamento para países da OECD (1990-2017)



Referências

- [1] Jacques Bertin. Semiology of graphics: Diagrams. *Networks, Maps*, 10(00690805.1987):10438353, 1983.
- [2] Wall Street Journal. Impact of vaccines throughout history. <https://flowingdata.com/2015/02/19/impact-of-vaccines-throughout-history/>, 2015. [Online].
- [3] Andy Kirk. *Data visualisation: a handbook for data driven design*. Sage, 2016.
- [4] Hans Rosling. The best stats you’ve ever seen. <https://www.youtube.com/watch?v=hVimVzgtD6w&t=3s>, 2007. [Online].