UNIVERSIDADE FEDERAL DO PARANÁ-UFPR

ESTATÍSTICA COMPUTACIONAL – CE 083

Conrado Mattei de Cabane Oliveira –GRR 20124694

Daniel Tyszka Júnior – GRR 20124667

Prof° Walmes Zeviani

Análise Exploratória dos dados brasileiros da desigualdade pela razão entre 20% mais ricos e 40% mais pobres e população economica-mente ativa dos anos 1991 e 2000.

12 de março de 2012

1. Introdução

O Brasil passou por um processo de mudança econômica durante a década de 90. Após o ex-presidente Collor subir ao poder, ele concedeu maiores liberdades de importação e exportação as empresas brasileiras. Foi a maior abertura dos portos desde a chegada da família real em 1808.

Além disso, no governo sucessório ao de Fernando Collor, houve a criação do Plano Real, uma tentativa, que hoje, verifica-se bem sucedida de controle inflacionário.

Demograficamente, o Brasil também passou por mudanças na década de 90. Desde o contínuo êxodo rural até o envelhecimento da população.

Desse modo, esse artigo tem como objetivo verificar se as mudanças econômicas e demográficas trouxeram avanços sociais. Ressalta-se, que esse é apenas um estudo pré-eliminar, e só será levado em conta, os dados de população economicamente ativa e a razão da renda entre os 20% mais ricos e os 40% mais pobres.

1.1. Dados

Para analisar a ocorrência de melhoras sociais no Brasil, utilizou-se o índice da de população economicamente ativa.

Conjuntamente, foi analisada a razão entre a renda dos 20% mais rico e 40% mais pobres. Os dados de ambas as variáveis se referem aos anos de 1991 e 2000.

Todos os dois conjuntos de dados foram retirados do Instituto de Pesquisa Econômica Aplicada (IPEA, www.ipeadata.gov.br), e estão a nível municipal.

Os dados das duas variáveis foram manipulados pelo *software R,* assim como a elaboração de gráficos e tabelas.

Por ser o estado com um terço do Produto Interno Bruto e ter 41 milhões de habitantes (dados IBGE), São Paulo foi escolhido com um estado a parte para se analisar juntamente com o Brasil essa evolução.

1.2. Variáveis

Renda - razão entre a renda dos 20% mais ricos e 40% mais pobres: Medido com frequência decenal pelo IPEA desde 1991 até 2000, é uma medida contínua, tem como objetivo medir o grau de desigualdade entre os indivíduos segundo a renda domiciliar *per capita*. Calcula a razão entre a renda média dos indivíduos pertencentes aos dois décimos mais ricos da população com a renda média dos indivíduos pertencentes aos quatro décimos mais pobres da população. A última atualização foi realizada em julho de 2011. Para maiores informações sobre a metodologia (www.undp.org.br).

População Economicamente Ativa (PEA) – Total: Medido com frequência decenal desde 1920 até 2000 pelo Instituto Brasileiro de Geografia e Estatística (IBGE), a população economica-

mente ativa mede o número de pessoas, ou seja, é uma variável discreta, dispostas a trabalhar nos últimos 12 meses. Por pessoas dispostas a trabalhar, o IBGE classifica as pessoas que exerceram trabalho remunerado, em dinheiro ou/e produtos ou mercadorias, inclusive as licenciadas, com remuneração por doença, com bolsa de estudo ou as sem emprego que trabalharam habitualmente 15 horas ou mais por semana numa atividade econômica, ajudando à pessoa com quem residiam ou à instituição de caridade, beneficente ou de cooperativismo, ou, ainda, como aprendizes ou estagiárias. Também foram consideradas nesta condição as pessoas de 10 anos ou mais de idade que não trabalharam nos doze meses anteriores à data de referência do Censo, mas que nos últimos dois meses tomaram alguma providência para encontrar trabalho. Elaboração IPEA, baseado no Censo Demográfico. No ano de 1940, os valores de Serra de Aimorés foram para o município de Carlos Chagas, por estar em litígio entre Minas Gerais e Espírito Santo. Para o ano de 1920 a fonte é Ministério da Agricultura, Indústria e Comércio. A sua última atualização ocorreu em fevereiro de 2011.

1. Análise das variáveis isoladamente;

1.1. Renda - razão entre a renda dos 20% mais ricos e 40% mais pobres.

1.1.1. Brasil

Como dito anteriormente, a razão entre a renda dos 20% mais ricos e 40% mais pobres é uma medida de desigualdade. Quanto maior a razão, maior a desigualdade na localidade. Se a razão é de 10, os 20% mais ricos têm uma renda média 10 vezes maior do que os 40% mais pobres.

O gráfico 1 é composto por dois histogramas, o da esquerda representa o ano de 1991 e o da direita representa o ano de 2000. Esses histogramas apresentam o comportamento da frequência em cada ano.

Gráfico 1

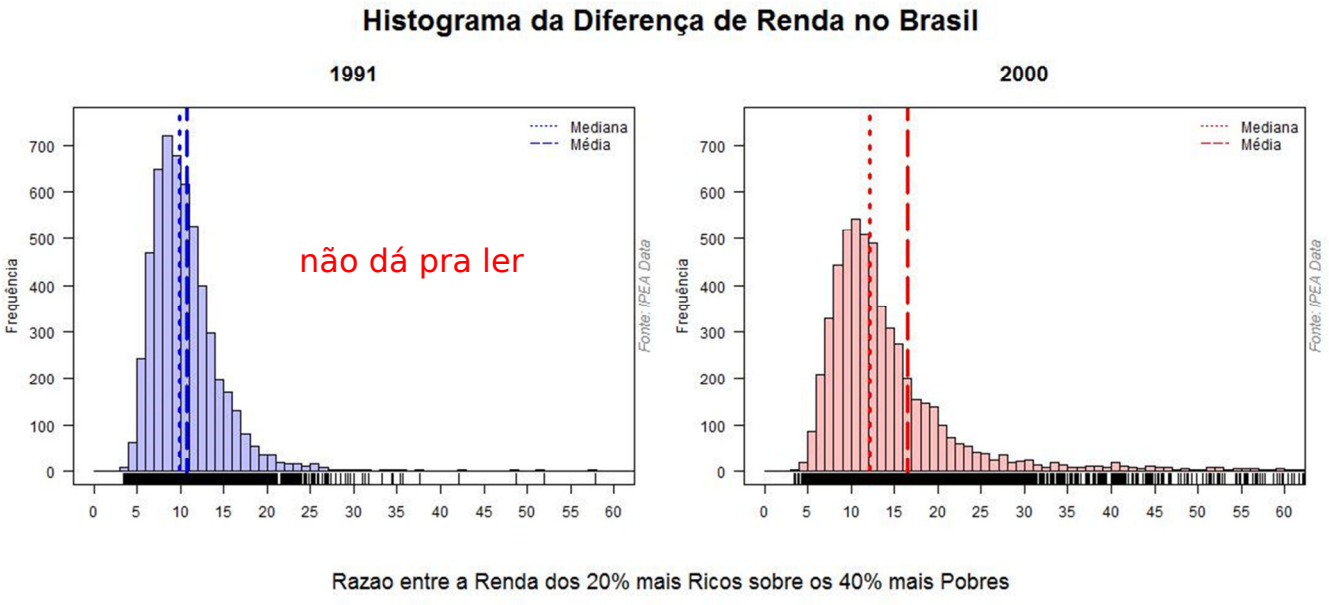


figura 1. ...

Primeiramente, observa-se que ambos os histogramas são assimétricos à direita, ou seja, suas médias, linhas tracejadas, são maiores que suas respectivas medianas, linhas pontilhadas. Quanto mais assimétrico à direita, mais justa é a renda do local de estudo.

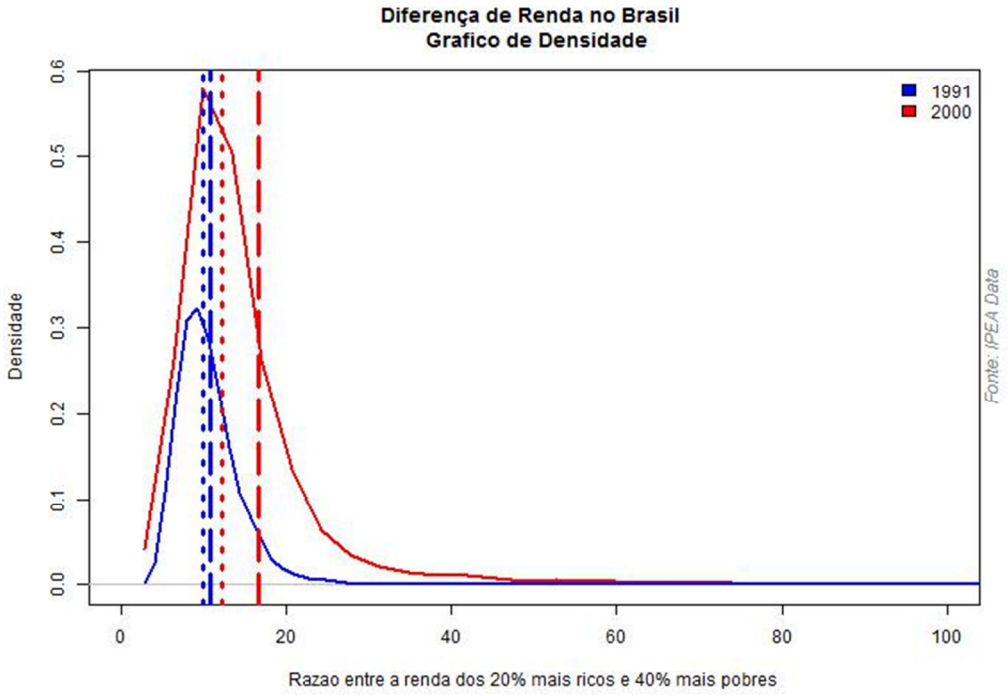
Em ambos os casos, a maior frequência ocorreu na razão 10, ou seja, os 20% mais ricos de 700 cidades em 1991, e 550 em 2000, tinham 10 vezes a renda média dos 40% mais pobres. É importante observar a quantidade de observações nas duas caudas desses histogramas. As marcações pretas, riscos nas verticais abaixo do eixo das abscissas, indicam uma observação dentro daquela frequência. Nota-se que, enquanto em 1991, após uma razão de 30 vezes, as marcações pretas rareiam, enquanto que no ano de 2000, elas continuam significativas. Ou seja, em 2000, houve mais cidades com uma razão de 30 vezes a renda do que no ano de 1991.

Os gráficos 2 e 3 fazem uma leitura complementar aos histogramas. O gráfico 2 é um gráfico de densidade. Nele, estão sobrepostos os valores de 1991 e 2000. Por definição, o eixo das ordenadas vai de 0 a 1. E podem ser traduzidos por uma porcentagem. No gráfico 2, podem-se entender melhor os histogramas anteriores.

Primeiramente, nota-se que a média, a linha tracejada, de 2000 está mais a direita do que a média de 1991. No início da década, o valor médio se aproximava da razão 10, enquanto que no final da década o valor ultrapassa a razão 15. Isso quer dizer que, entre 1991 e 2000, o Brasil ficou mais desigual. Os 20% mais ricos aumentaram mais a sua renda do que os 40% mais pobres.

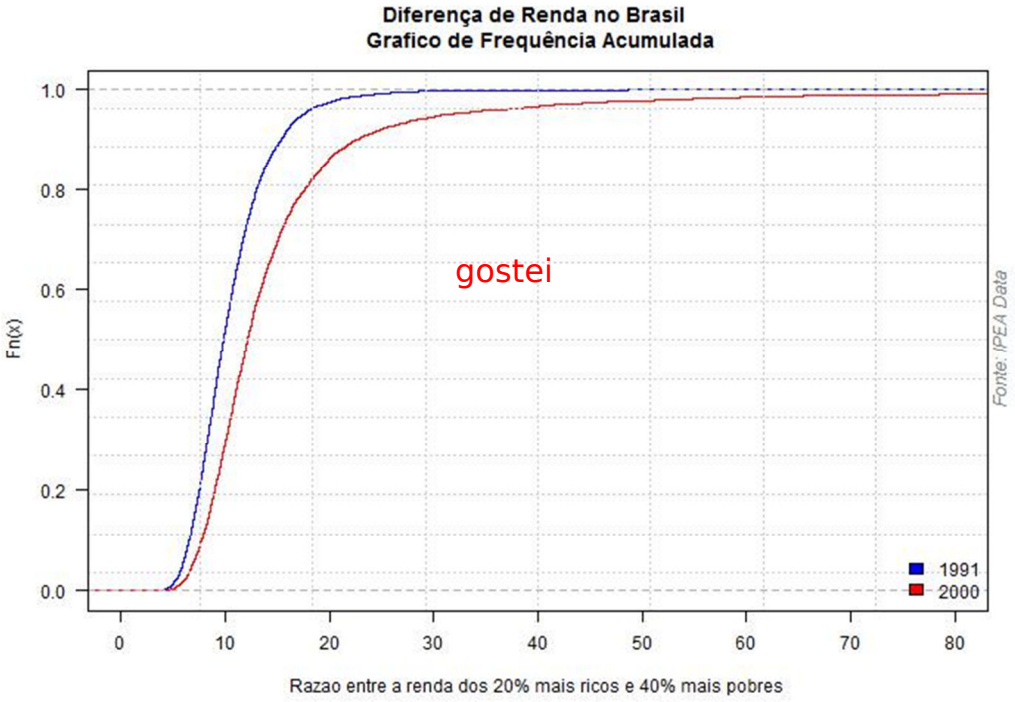
Além disso, após a razão 10, o gráfico de 1991 aproxima-se de zero, ou seja, baixa frequência nos valores subsequentes. Enquanto que o gráfico de 2000 tem uma cauda mais estendido, tendo valores com razões de concentração mais altos, corroborando os riscos pretos dos histogramas.

Gráfico 2



O gráfico 3, ajuda a visualizar a situação da cauda. O gráfico a seguir é um gráfico de frequência acumulada de ambos os anos.

Gráfico 3

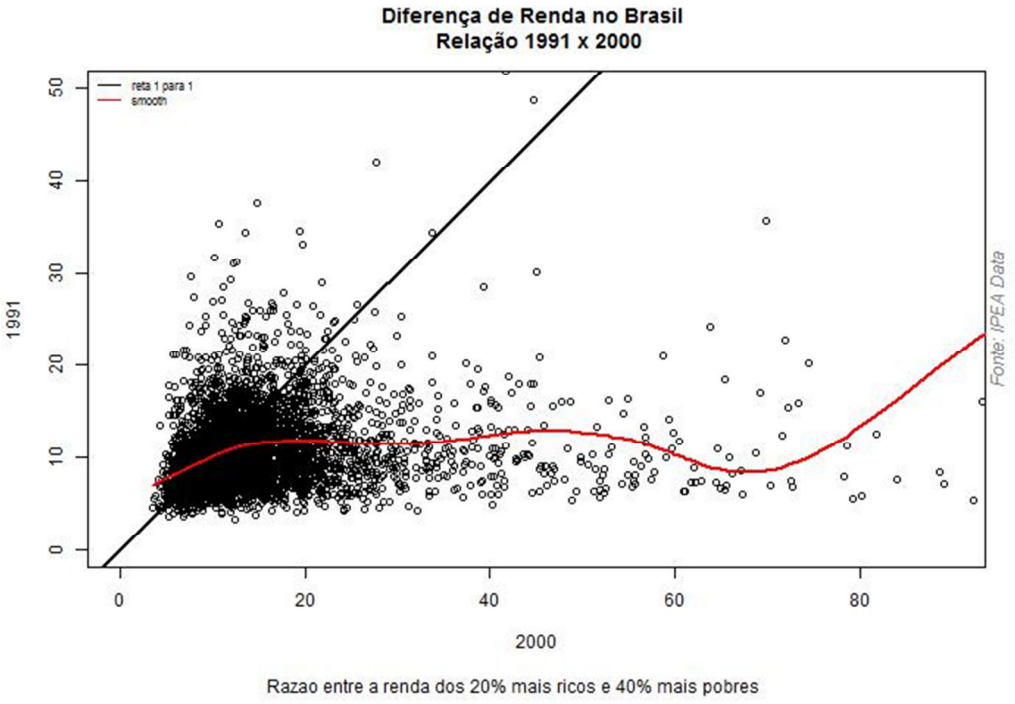


Um gráfico de frequência acumulada, como o nome sugere, conforme o eixo x se dirige à direita, o eixo y soma a frequência até o momento. Ele acumula a frequência. Assim, os valores no eixo das ordenadas vão de zero (nenhuma frequência somada) a 1 (100% da frequência somada).

Observa-se como o gráfico de 1991 engloba o de 2000. Isso significa que as menores razões, se acumulam antes que em 2000. Com a razão em 20, o ano de 1991 tem quase 100% dos valores, ou seja, quase 100% dos valores estão abaixo dos valores da razão 20. Já no gráfico de 2000, na razão 20 falta ainda 10% da frequência (ou dos municípios brasileiros).

Junto com esses três gráficos, o gráfico 4 ajuda a corroborar como a renda se concentrou na passagem da década de 2000.

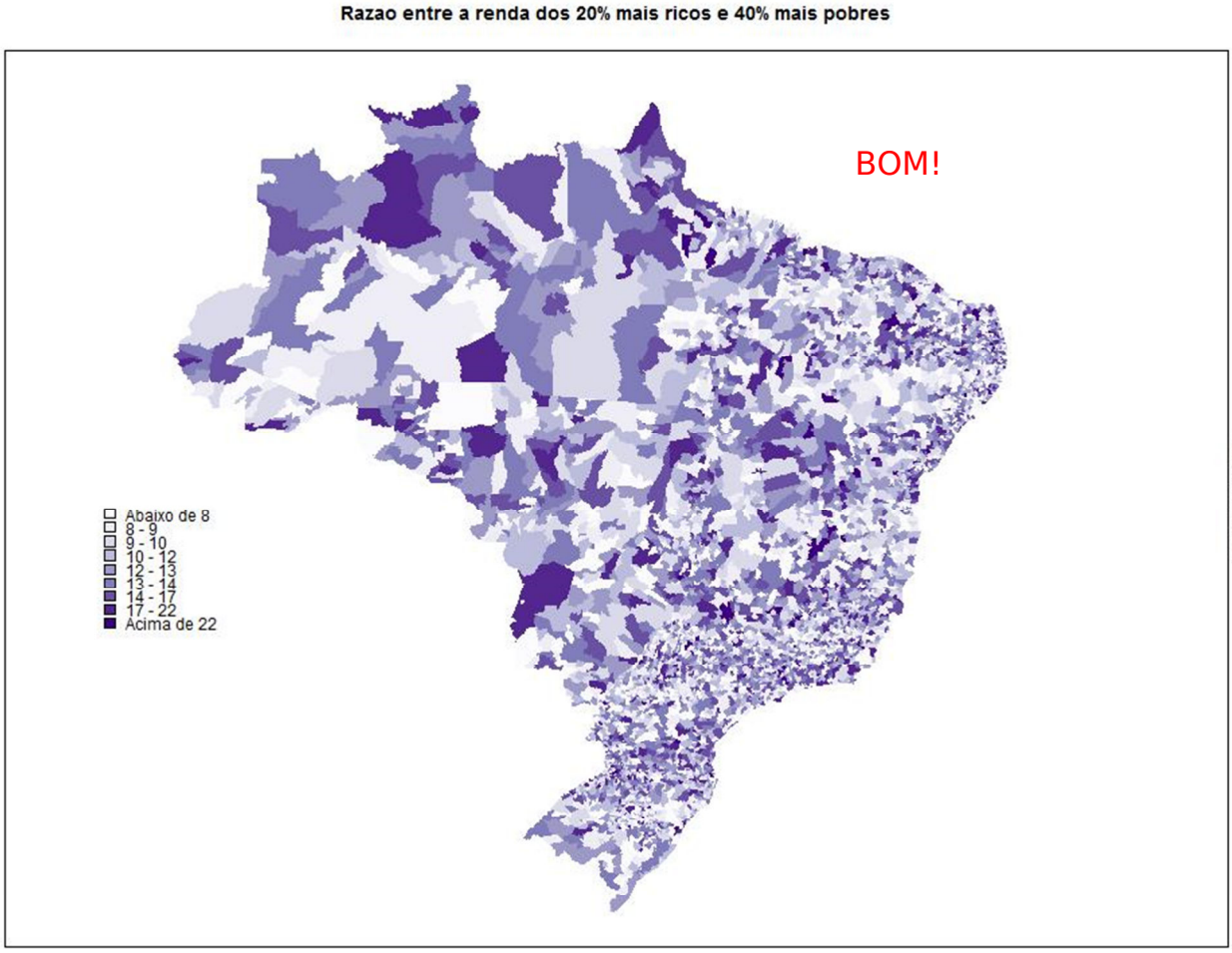
Gráfico 4



O gráfico 4 apresenta um gráfico de dispersão, no qual cada ponto é um município brasileiro. Os pontos acima da linha preta indicam os municípios que tem menos concentração de renda no ano de 2000 do que no ano de 1991. Consequentemente, os pontos abaixo da linha preta indicam o movimento inverso, o de concentração de renda.

Percebe-se uma grande quantidade abaixo da linha, principalmente na nuvem principal e ao longo dela. A linha vermelha ajuda a indicar essa tendência. A linha vermelha, chamada de *smooth* (suave, suavização em inglês), indica para que lado os pontos seguem. Infelizmente, no presente caso, a linha segue para a direita, indicando uma concentração de renda maior no ano de 2000 do que no ano de 1991.

Gráfico 5



O gráfico 5 apresenta um mapa do Brasil de 2000, com os limites de cada município colorido com seu respectivo índice de diferença entre os ricos e pobres. Quanto mais escura a cor roxa, mais desigualdade no respectivo município.

Observa-se que é raro encontrar pontos com as tonalidades mais escuras de roxo nos estados do sul e sudeste (com razões entre 17 e 22 e acima de 22). Porém, conforme avança para o norte e nordeste, inclusive em alguns pontos do centro-oeste, os pontos mais roxos vão ficando mais evidentes. É claro, que há cidades com baixa concentração nesses estados. Isso tudo corrobora os pontos dos gráficos anteriores, nos quais os *blox plots* que é possível encontrar cidades com baixa concentração de renda no norte e nordeste (barra inferior do *box plot*), menor concentração de renda nos estados do sul, sudeste e centro-oeste (barras com razões médias menores do que em outras regiões).

1.1.2. Estados

O gráfico 6, apresenta a razão da renda dos mais ricos em relação aos mais pobres nos *box plots* de cada estado, com os valores de 2000.

Um *box plot* apresenta os cinco valores de posição do sumário estatístico das observações, valor inferior, valor superior, primeiro quartil, segundo quartil mediana e terceiro quartil.

Observa-se que os estados do sul, sudeste, a mediana, ou seja, onde estão 50% dos dados se encontram perto da razão 10. Esses estados, também apresentam baixa dispersão, com pequenas caixas e baixa amplitude.

Enquanto isso, os estados do norte e nordeste apresentam uma mediana mais alta, e caixas maiores, apresentando uma maior dispersão. Destaque, neste caso negativo, para os estados do Acre, Tocantins e Amazonas, que apresentam as maiores diferenças entre pobres e ricos.

Gráfico 6

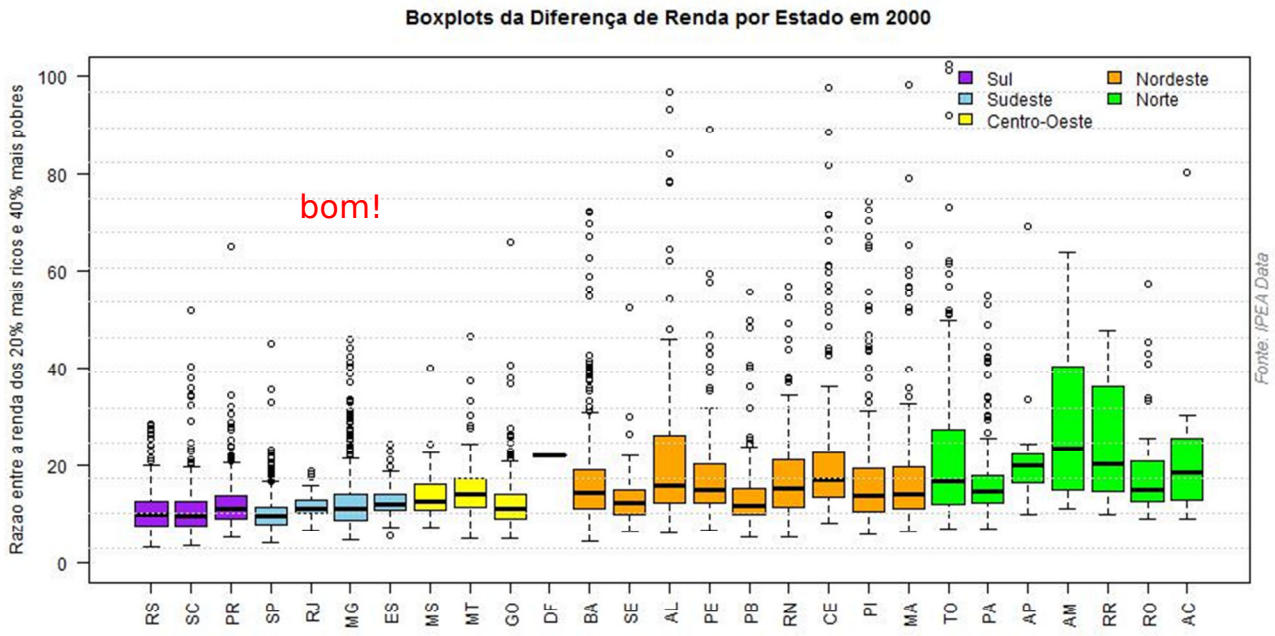


figura. ...

Junto ao gráfico 6, a tabela 1 nos ajuda a entender a diferença de renda em cada estado.

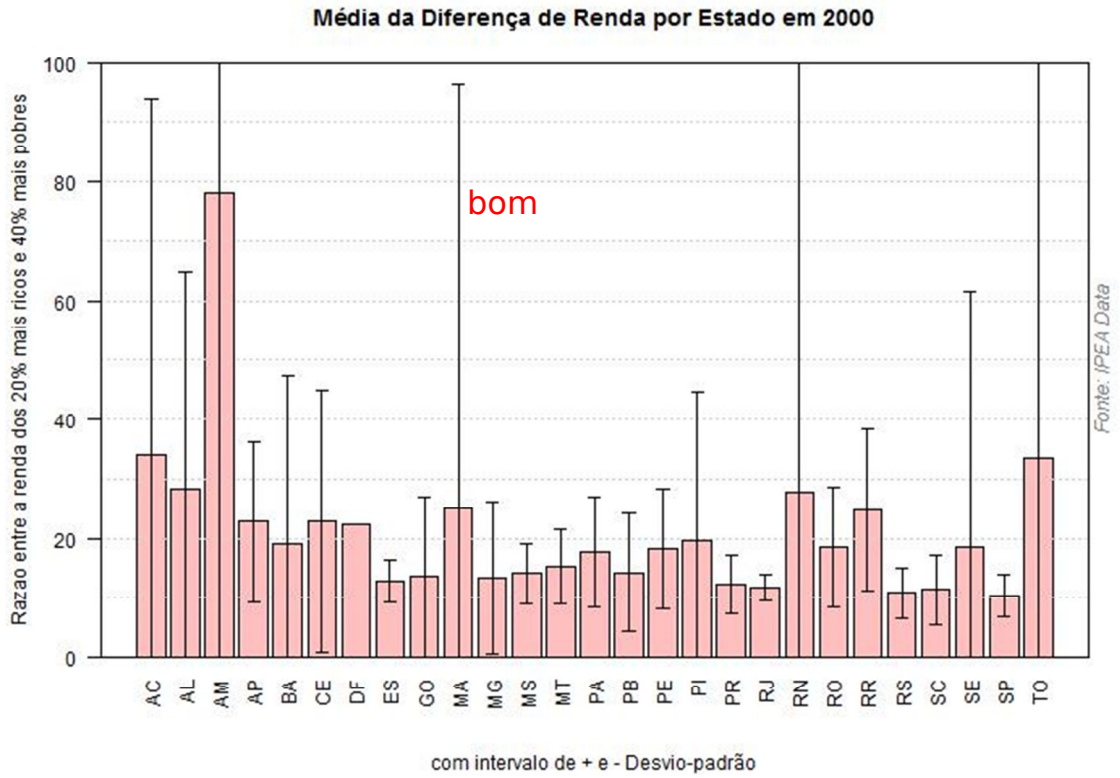
Tabela 1 legenda

	Media	Mediana	Minimo	Maximo	Quartil1	Quartil3	Amplitude	Desvio-Padrao	Desvio Absoluto Medio	n-Amostra
AC	34	19	9	293,19	13	26	284	60	20	22
AL	28	16	6	267	13	26	261	37	16	100
AM	78	24	11	1820,29	15	40,5	1.809	241	62	58
AP	23	20	10,1	69,28	17	22	59,18	13	7	16
BA	19	14	5	517	11	19	512	28	8	412
CE	23	17,21	8	237	14	23	229	22	10	184
DF	22	22	22	22	22	22	0	NA	0	1
ES	13	12	6	25	11	14	19	3	2	77
GO	14	11	5	201	9	14	195	14	4	242
MA	25	14	7	1.006	11	20	999	71	14	214
MG	13	11	4,94	269	9	14	264	13	4	851
MS	14	13	7	40	11	16	33	5	3	77
MT	15	14	5,16	46,61	12	17	41,45	6	4	126
PA	18	15	7	54,9	12	18	48	9	6	143
PB	14	12	6	120	10	16	114	10	5	223
PE	18	15	7	89	12	20	82	10	6	182
PI	20	14	6	316	11	19	310	25	9	218
PR	12	11	5	65	9	14	59	5	3	399
RJ	12	11,2	7	19	10	13	12	2	2	91
RN	28	15,34	5	1.238	12	21	1.233	97	16	164
RO	19	15	9	57	13	20,91	48,16	10	6	52
RR	25	20	10	48	14,69	36	38	14	10	12
RS	11	10	3	29	8	13	25,13	4	3	467
SC	11	10	4	52	8	13	48	6	4	293
SE	19	13	6	381	10	15	375	43	8	75
SP	10	10	4	45	8	12	41	4	2	645
TO	34	17	7	919	12,25	28	912	83	21	138

O gráfico 7 é um complemento dos *box plots* acima. A diferença desse gráfico é que além das barras com as médias, o gráfico informa o desvio-padrão, ou seja, o quanto que na média as observações variam da média.

Como no Distrito Federal só há um dado, o de Brasília, os dados do sumário estatístico são iguais, e não há desvio padrão ou amplitude. Por essa razão, não há *box plot,* apenas uma barra.

Gráfico 7



O gráfico 7 apresenta de forma mais clara a média da diferença da razão da renda no ano de 2000. O estado do Amazonas apresenta, disparado, a maior diferença, com uma razão de quase 80 vezes, ou seja, no estado os 20% mais ricos tem uma renda médias 80 vezes maiores que os 40% mais pobres.

Destaque, positivo, para os estados do sul e sudeste, com uma diferença próxima da razão 10, seguida pelos estados do centro-oeste, que exceto o distrito federal, não passaram a razão de 20. Além disso, os estados dessas três regiões apresentam um baixo desvio-padrão. Ou seja, dentro dos estados, a probabilidade de haver grandes diferenças de rendas entre os municípios é pequena.

1.1.3. São Paulo

Como foi visto anteriormente, os estados do sul, sudeste e centro-oeste possuem as menores diferenças entre os 20% mais ricos em relação aos 40% mais pobres.

O gráfico 8 mostra em um histograma como o dado da desigualdade se observa no Estado de São Paulo a partir das suas frequências no ano de 2000.

Gráfico 8

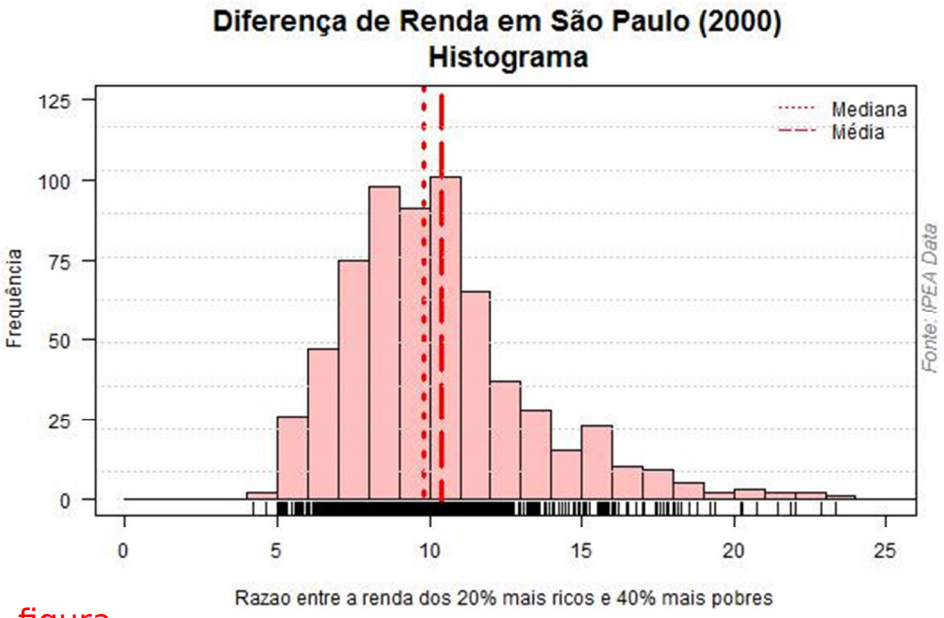
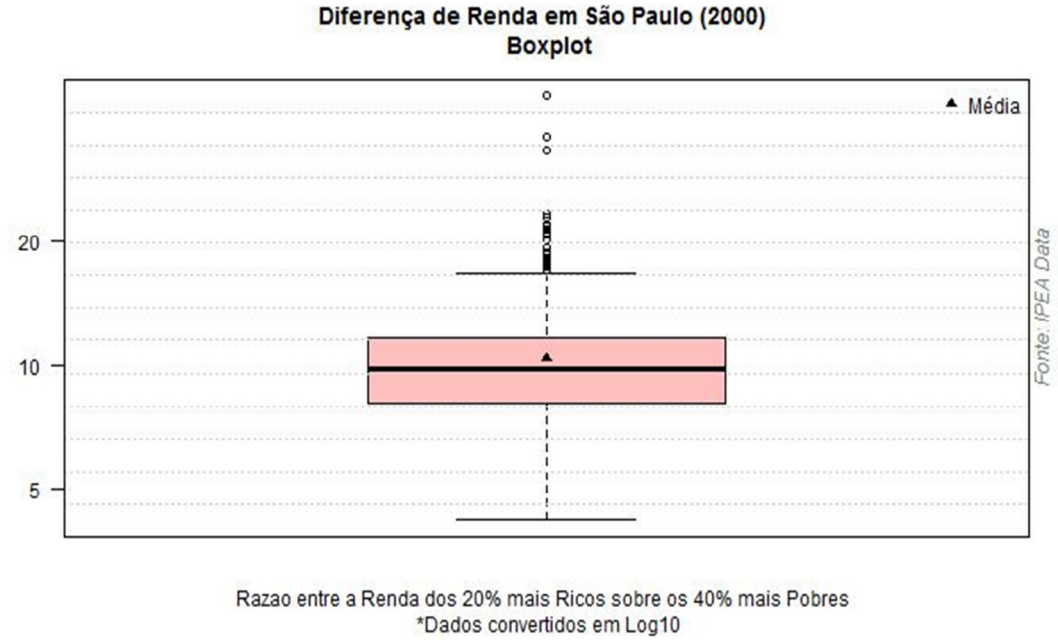


figura. ...

O gráfico nacional também tinha dado com uma assimetria à direita, porém, essa assimetria é mais suave no estado de São Paulo. A distância entre a média e a mediana está menor do que o gráfico nacional. Ou seja, os dados estão mais concentrados no estado de São Paulo. Isso pode ser verificado pela cauda a direita do histograma acima. No caso brasileiro, ela era mais estendida, com valores inclusive acima da razão 25. Em São Paulo, a partir da razão 20, as observações começam a rarear. Ou seja, no estado há poucos municípios onde os 20% mais ricos possuem uma renda superior a 20 vezes a renda dos 40% mais pobres. No histograma, verifica-se que há aproximadamente 100 municípios que a razão da renda chega a 10 vezes.

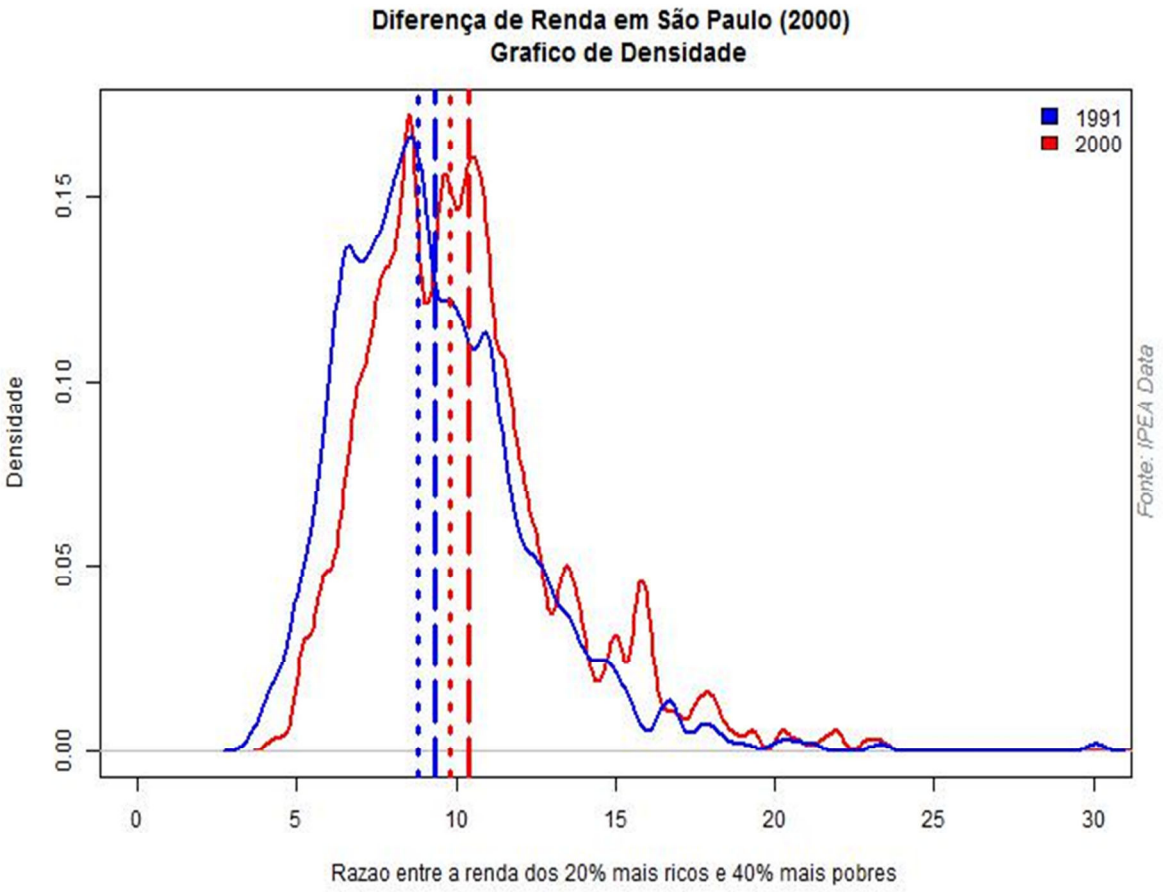
O *box plot* do gráfico 9 é um complemento do histograma . Visualiza-se que a maioria dos dados se encontram um poço abaixo da razão 5 e um pouco abaixo da razão 20 (diferença entre a barra superior e a barra inferior). Conclui-se que é normal um município de São Paulo ter essa diferença de renda.

Gráfico 9



Infelizmente, ocorreu em São Paulo o mesmo fenômeno que ocorreu nacionalmente. Entre 1991 e 2000, a diferença de renda entre os maios ricos e os mais pobres aumentou. Como pode ser visto no gráfico 10, a média e a mediana do ano 2000 estão mais a direita do os respectivos valores de 1991. Isso quer dizer que ao longo da década de 90, os 20% mais ricos concentraram mais renda do que os 40% mais pobres.

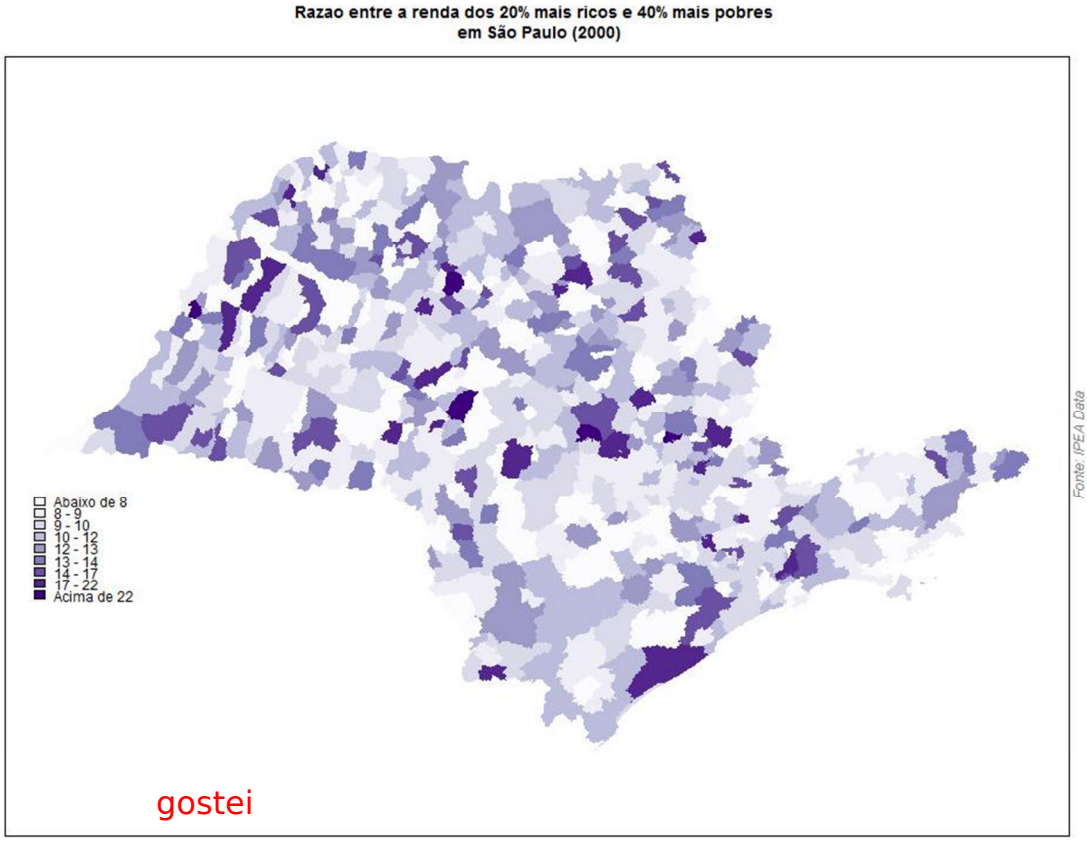
Gráfico 10



O gráfico abaixo mostra o mapa do estado de São Paulo de acordo com a diferença de renda de cada município. Quanto mais escuro for o roxo, maior a diferença de renda naquela localidade.

Observa-se que no centro de São Paulo, na região da capital e de Campinas, há um clarão, indicando que naquela região há uma razão inferior a 12 vezes. Perto da região de Minas Gerais também há uma região clara.

Gráfico 11


Razao entre a renda dos 20% mais ricos e 40% mais pobres
em São Paulo (2000)
Abaixo de 8
8 - 9
9 - 10
10 - 12
12 - 13
13 - 14
14 - 17
17 - 22
Acima de 22
Fonte: IPEA Data
gostei

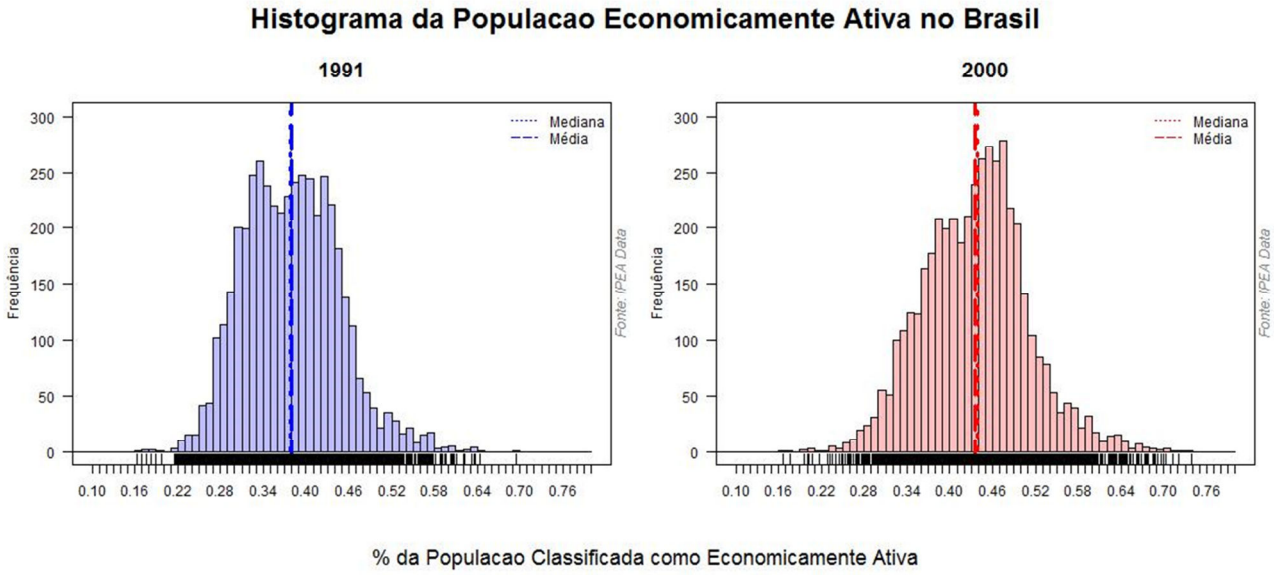
1.2. População economicamente ativa (PEA)

1.2.1. Brasil

Originalmente, os valores do PEA encontram-se em valores discretos, pelo número de pessoas. Porém, para tentar minimizar possíveis distorções por questões de tamanho populacional (O PEA do município de São Paulo é maior que a população total de qualquer cidade brasileira), dividimos o PEA pela população total, criando uma nova variável, agora contínua, que varia entre 0 a 1, indicando o percentual da população economicamente ativa dentro daquele município. Ressalta-se, que ambos os valores têm a mesma fonte, o IPEA.

O gráfico 12 apresenta dois histogramas, o da esquerda referente ao ano de 1991 e o da direita, referente ao ano 2000.

Gráfico **1**

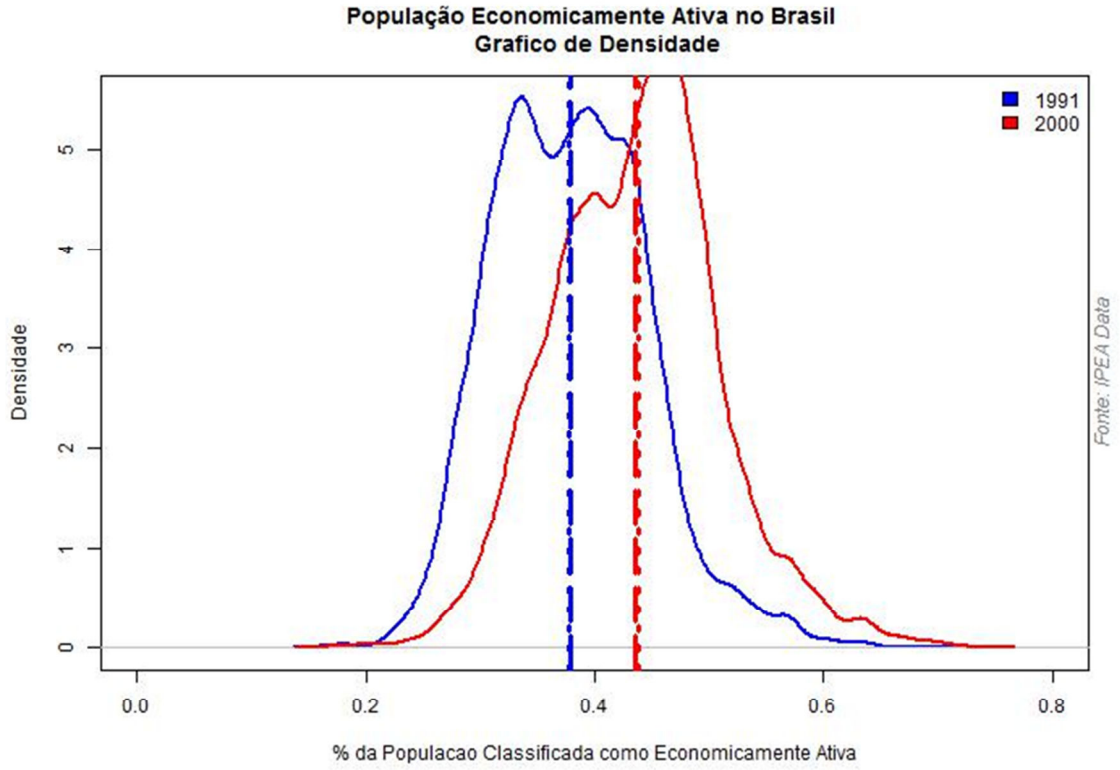


Ambos têm uma forma semelhante, sendo mais concentrados nos valores medianos, e com caudas em ambos os lados. Vê-se que nos dois anos pesquisados, as médias encontram-se próximos à mediana. Os dois histogramas assemelham-se a uma curva normal.

Porém, observa-se certa evolução do PEA na década de 90. Os gráficos a seguir nos mostram de forma mais clara.

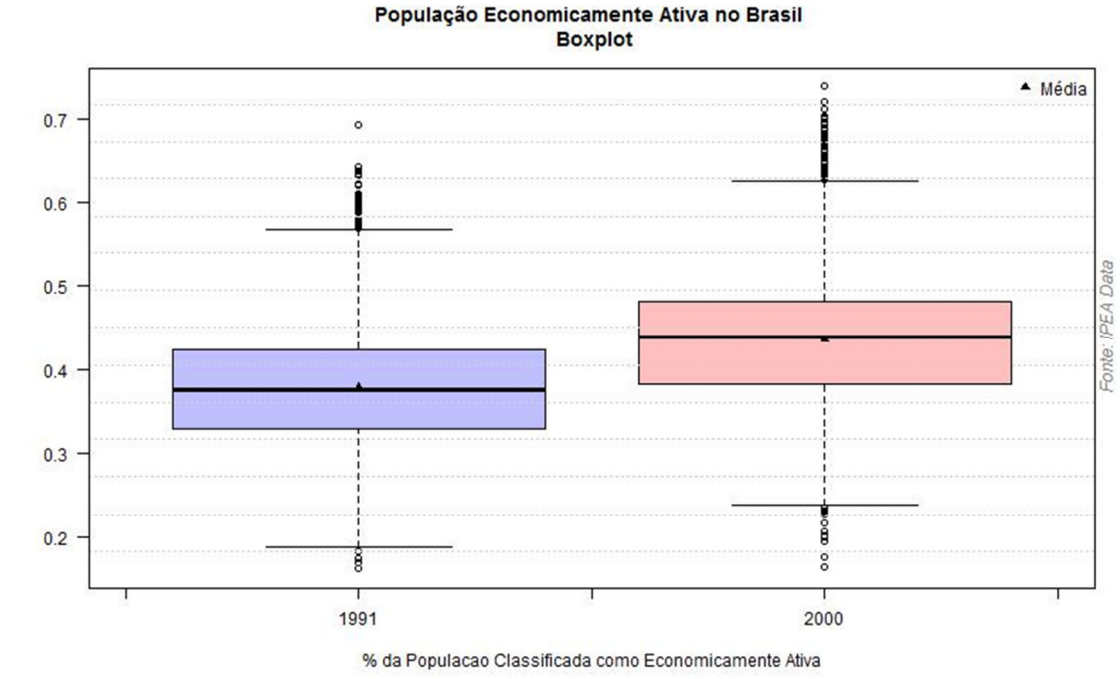
O gráfico 13, é um gráfico de distribuição da frequência de ambos os anos, um sobreposto ao outro. Nele, se observa um deslocamento do gráfico para a direita do gráfico do ano de 2000 em relação ao gráfico de 1991, ou seja, houve um aumento nos valores das frequências nos maiores valores do PEA. A linha da média confirma essa análise anterior. A média de 2000 está mais a esquerda que a média de 1991. Ou seja, no ano de 2000, havia mais pessoas dispostas a trabalhar do que no ano de 1991, isso, diga-se, proporcionalmente.

Gráfico 13



Os *box plots*, do gráficos 14, ajuda nessa análise.

Gráfico 14



Observa-se que a mediana, a barra dentro da caixa, e que divide 50% dos valores para cima e para baixo desse valor, no ano de 2000 está acima do que do ano de 1991. A barra inferior do ano de 2000 também está acima do ano de 1991, assim como as outras três estatísticas de posição, primeiro quartil, terceiro quartil e barra superior. Verifica-se a presença de *outliers*, tanto em cima, quanto embaixo nos dois *box plots.*

Dentro dos *outliers*, surpreende o fato de haver municípios com valores do PEA equivalentes a 20% da população, ou seja, apenas 1 quinto da população está disposta a trabalhar.

A tabela 2 é um complemento do *box plot* acima.

Tabela 2 legenda

Variável	1991	2000
Media	0.379	0.436
Mediana	0.377	0.439
Minimo	0.163	0.166
Maximo	0.694	0.739
Quartil1	0.329	0.384
Quartil3	0.425	0.481
Amplitude	0.531	0.574
Desvio-Padrao	0.067	0.073
Desvio Absoluto Medio	0.054	0.058
n-Amostra	4491	4491

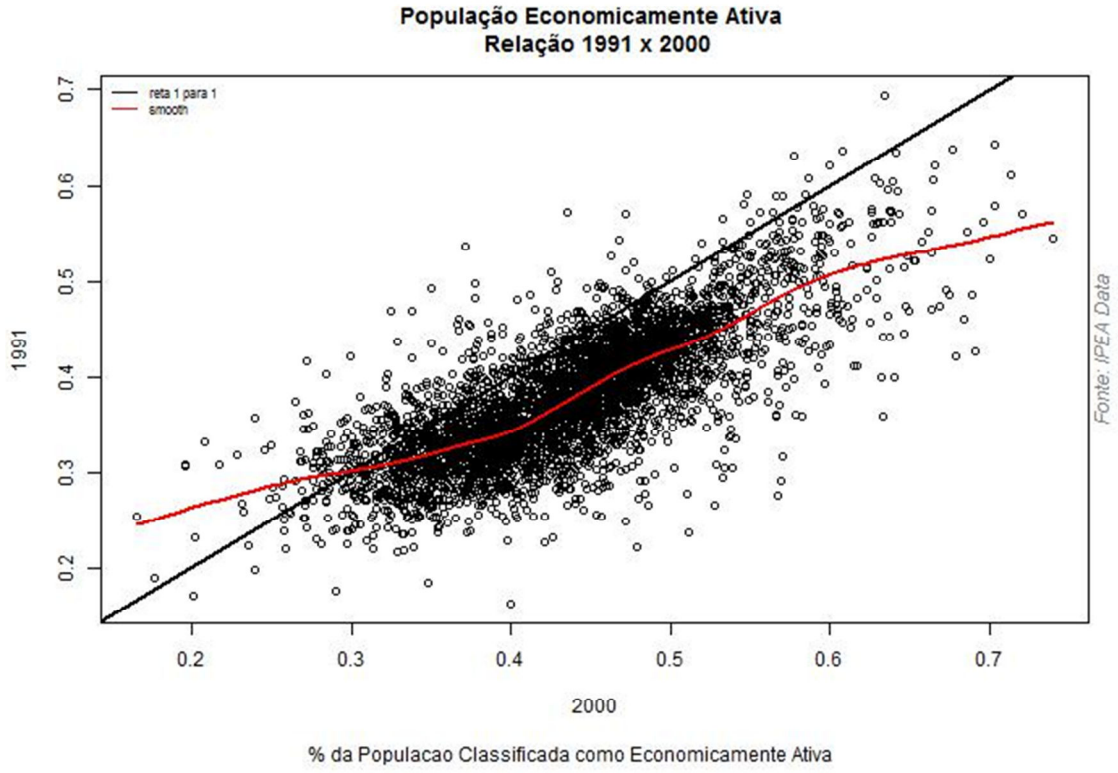
A tabela 2 mostra numericamente a evolução demonstrada nos *box plot.* Verifica-se a evolução das médias e medias de 37% cada uma no ano de 1991 para 43% no período de 2000.

O número mínimo se manteve estável em 16%, ou seja, há municípios com apenas 16% da população disposta a trabalhar, enquanto que a o máximo passou para 73%, ou seja, é possível encontrar municípios com até 73% disposta a trabhalhar.

O gráfico 15, o gráfico de dispersão demonstra um pouco mais dessa evolução, entre o PEA de 1991 e o PEA de 2000.

O gráfico 15 representa a dispersão de cada município, de acordo com o cruzamento dos valores de cada ano. Valores abaixo da linha preta indicam que o PEA daquele município é maior no ano 2000 do que em 1991, ou seja, a mais pessoas dispostas a trabalhar naquele município proporcionalmente. Já os valores acima da linha indicam o inverso.

Gráfico 15.

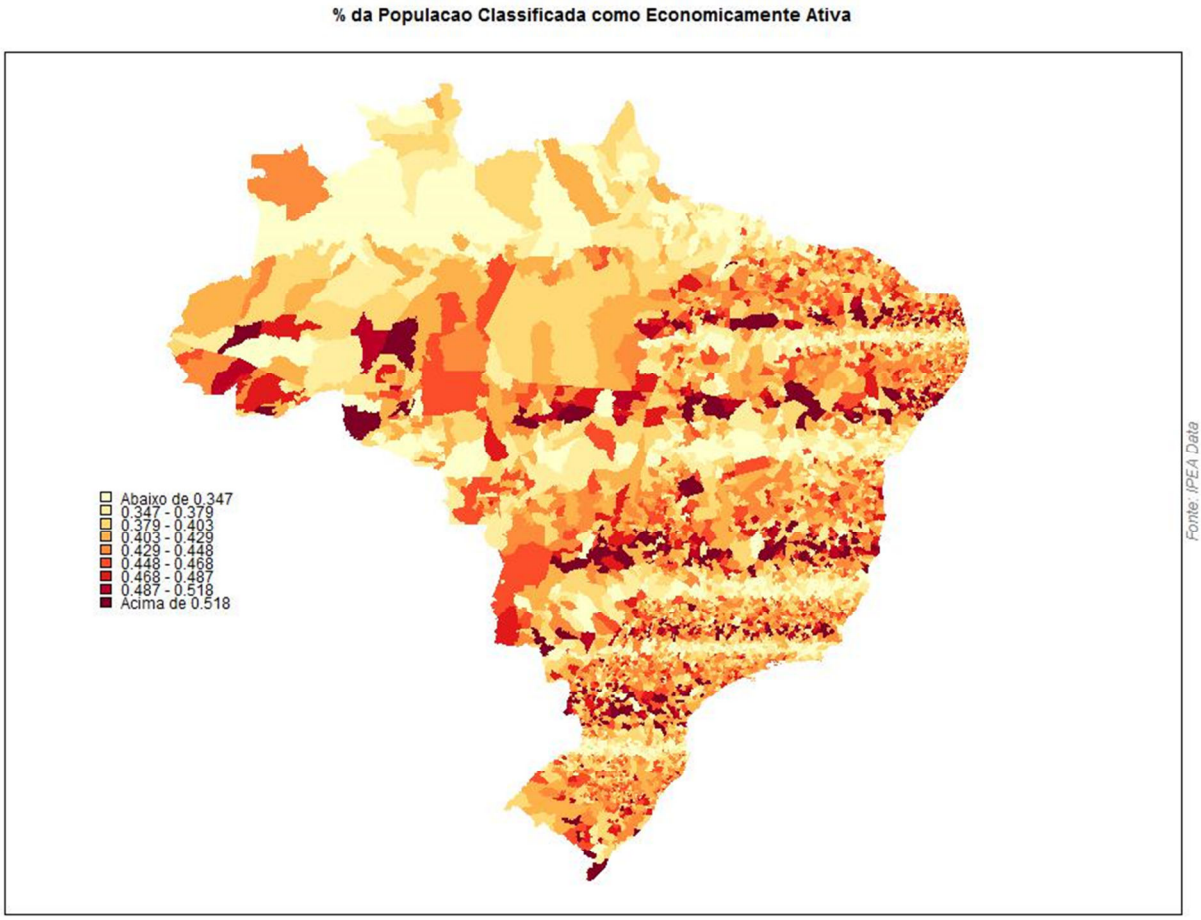


Nesse caso, é visível que há mais valores abaixo da linha do que a cima, corroborando os gráficos anteriores. A linha vermelha, a linha de *smooth,* tende a ir abaixo da linha preta, 'trazendo' os dados junto com ela. Observa-se que no ponto em que a linha *smooth*, encontra-se acima da linha preta, há também mais valores acima da linha preta. Ou seja, a linha *smooth* indica uma tendência. Ou seja, se não houvesse os pontos, apenas as linhas, poderíamos nos guiar pela linha vermelha que teríamos uma grande probabilidade de acerto da posição do ponto naquela região, acima ou abaixo da linha preta.

O gráfico 16 é um mapa do Brasil, com cada município representado com a cor indicando seu respectivo índice PEA do ano de 2000. Quanto mais escuro for o amarelo, maior o PEA, e consequentemente, quanto mais claro o amarelo, menor o PEA.

O mapa corrobora as analises anteriores. No norte, na região do Amapá e do Amazonas, menores índices de PEA. A região norte está mais clara do que as demais regiões. É importante notar algumas regiões mais escuras no nordeste, e que há cidades no norte com mais de 50% da população disposta a trabalhar (os *outliers* dos *box plots*). No Paraná e no centro do Brasil encontram-se regiões com bastantes cidades escuras. Há faixas claras saindo do litoral, porém não passam de coincidência.

Gráfico 16



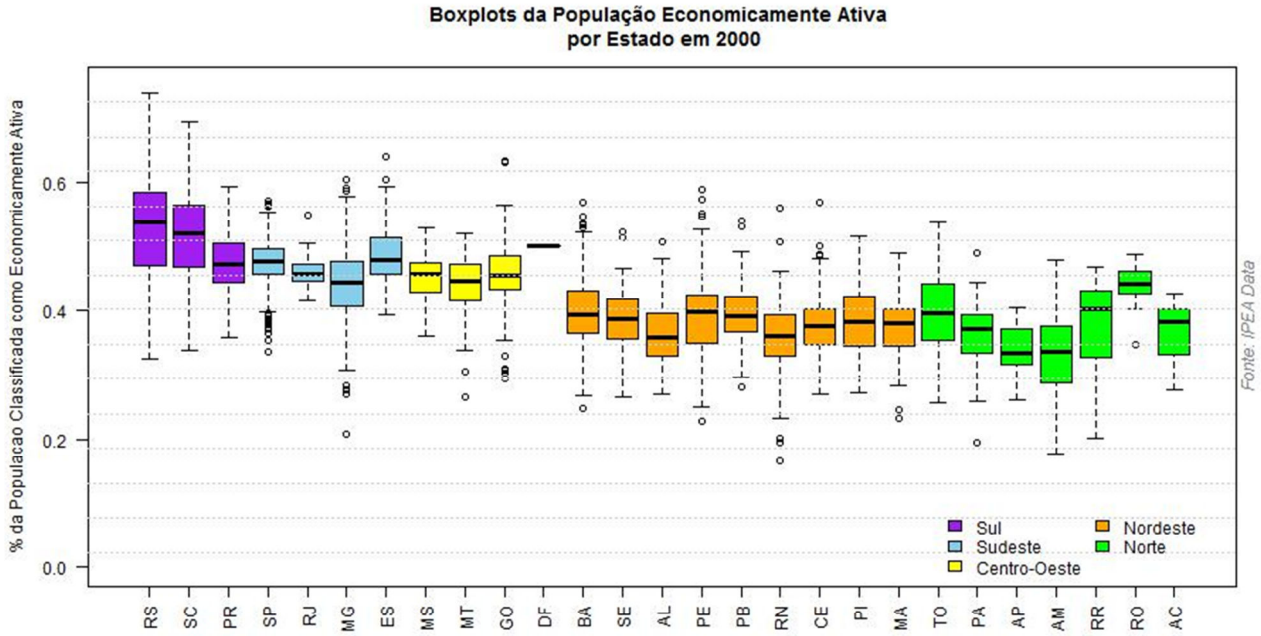
1.2.2. Estados

O gráfico 17 e 18 ajudam a entender a distribuição e dispersão do PEA em cada estado.

Pelos *box plots* do ano 2000, destaca-se o Rio Grande do Sul e Santa Catarina com os maiores índices percentuais de PEA. Paraná e os estados do sudeste apresentam valores também acima da média. Destaque negativo para o Amazonas e Roraima, estados com barras inferiores próximos ou além de 0,2%, ou seja, o equivalente a 1 quinto da população não está disposta a trabalhar.

Dentre os estados nortistas, destaque positivo para Rondônia, que tem os maiores valores da região, equivalente ao estado do Rio de Janeiro.

Gráfico 17



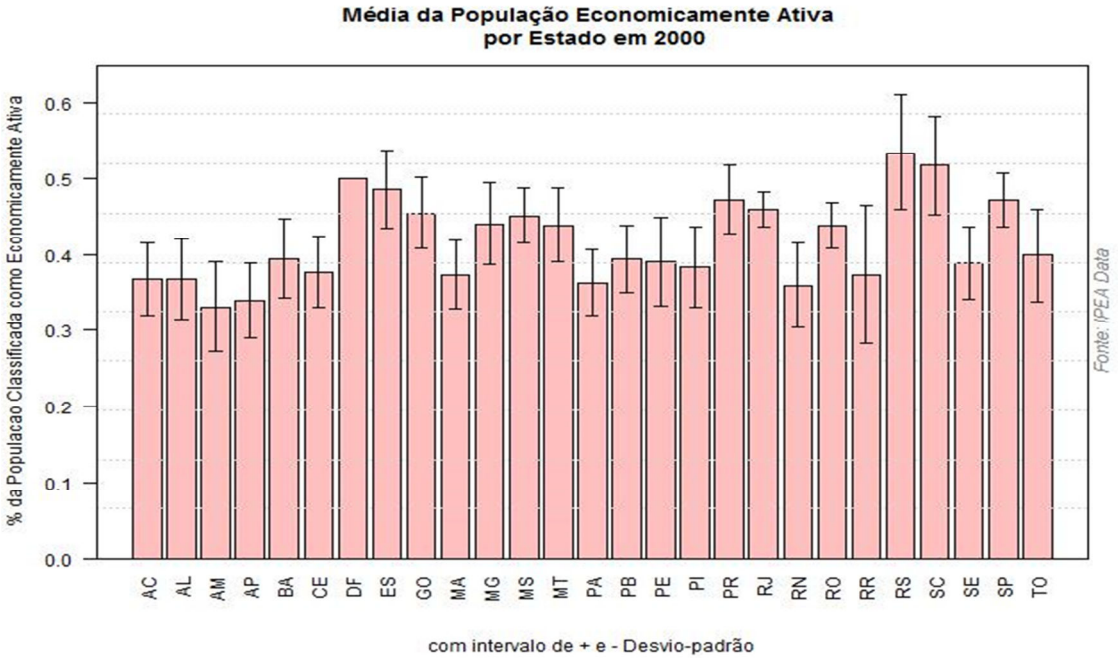
A tabela 3 auxilia na visualização dos dados por estado.

Tabela 3

	Media	Mediana	Minimo	Maximo	Quartil1	Quartil3	Amplitude	Desvio Padrão	Desvio Absoluto Médio	n-Amostra
AC	0,368	0,383	0,276	0,426	0,333	0,403	0,150	0,047	0,036	12
AL	0,367	0,358	0,270	0,508	0,330	0,395	0,238	0,053	0,042	97
AM	0,332	0,337	0,176	0,479	0,288	0,375	0,303	0,060	0,048	62
AP	0,340	0,333	0,262	0,405	0,316	0,370	0,143	0,049	0,039	9
BA	0,395	0,394	0,249	0,571	0,365	0,431	0,322	0,052	0,041	415
CE	0,378	0,377	0,271	0,569	0,347	0,401	0,298	0,047	0,036	178
DF	0,501	0,501	0,501	0,501	0,501	0,501	0,000	NA	0,000	1
ES	0,487	0,479	0,393	0,640	0,457	0,515	0,247	0,051	0,039	67
GO	0,456	0,455	0,294	0,633	0,432	0,486	0,340	0,047	0,034	211
MA	0,374	0,379	0,232	0,490	0,344	0,403	0,258	0,044	0,035	136
MG	0,441	0,444	0,208	0,604	0,406	0,477	0,396	0,054	0,043	723
MS	0,451	0,457	0,361	0,530	0,428	0,476	0,169	0,036	0,028	72
MT	0,439	0,446	0,265	0,521	0,414	0,474	0,257	0,048	0,036	95
PA	0,363	0,371	0,196	0,491	0,334	0,393	0,295	0,044	0,033	105
PB	0,394	0,392	0,281	0,540	0,367	0,421	0,259	0,044	0,034	171
PE	0,391	0,399	0,228	0,589	0,349	0,425	0,361	0,059	0,045	168
PI	0,384	0,382	0,272	0,516	0,345	0,422	0,244	0,053	0,043	118
PR	0,473	0,473	0,358	0,594	0,444	0,505	0,236	0,045	0,036	323
RJ	0,460	0,457	0,415	0,550	0,446	0,473	0,135	0,024	0,017	70
RN	0,359	0,360	0,166	0,561	0,328	0,393	0,395	0,056	0,042	152
RO	0,439	0,443	0,347	0,487	0,426	0,462	0,140	0,031	0,023	23
RR	0,374	0,403	0,201	0,469	0,327	0,431	0,268	0,091	0,062	8
RS	0,535	0,539	0,324	0,739	0,471	0,585	0,415	0,076	0,063	333
SC	0,518	0,522	0,338	0,696	0,469	0,565	0,358	0,065	0,053	217
SE	0,389	0,387	0,265	0,524	0,355	0,416	0,259	0,048	0,037	74
SP	0,473	0,477	0,336	0,571	0,456	0,496	0,235	0,035	0,027	572
TO	0,399	0,395	0,256	0,538	0,354	0,442	0,282	0,061	0,049	79

O gráfico 18 ajuda na análise do gráfico 17. Percebe-se o Rio Grande do Sul na ponteira com mais de 50% da população disposta a trabalhar no ano de 2000 seguido por Santa Catarina e Distrito Federal. Já o estado do Amazonas e Amapá, na outra ponta da tabela, com um pouco mais de 30% da população disposta a trabalhar.

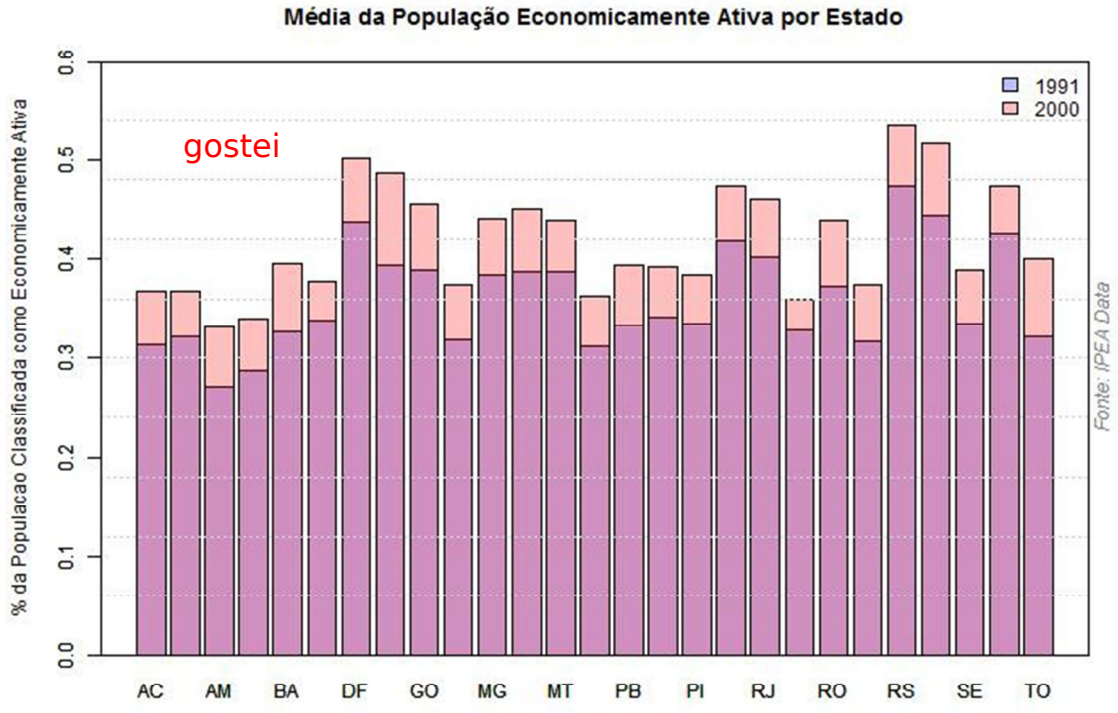
Gráfico 18



Nota-se que nesse caso, devido aos valores terem se convergido muito para o centro do gráfico, temos desvios-padrões menos que nas variáveis anteriores.

Se os gráficos anteriores nos dizem a distribuição e dispersão, o gráfico 19 ajuda a entender a evolução do PEA de 1991 ao ano 2000 em cada estado.

Gráfico 19



O gráfico 19 nos indica que houve um aumento no índice do PEA em relação ao total da população em todos os estados brasileiros. O estado que teve o menor aumento foi o estado do Rio

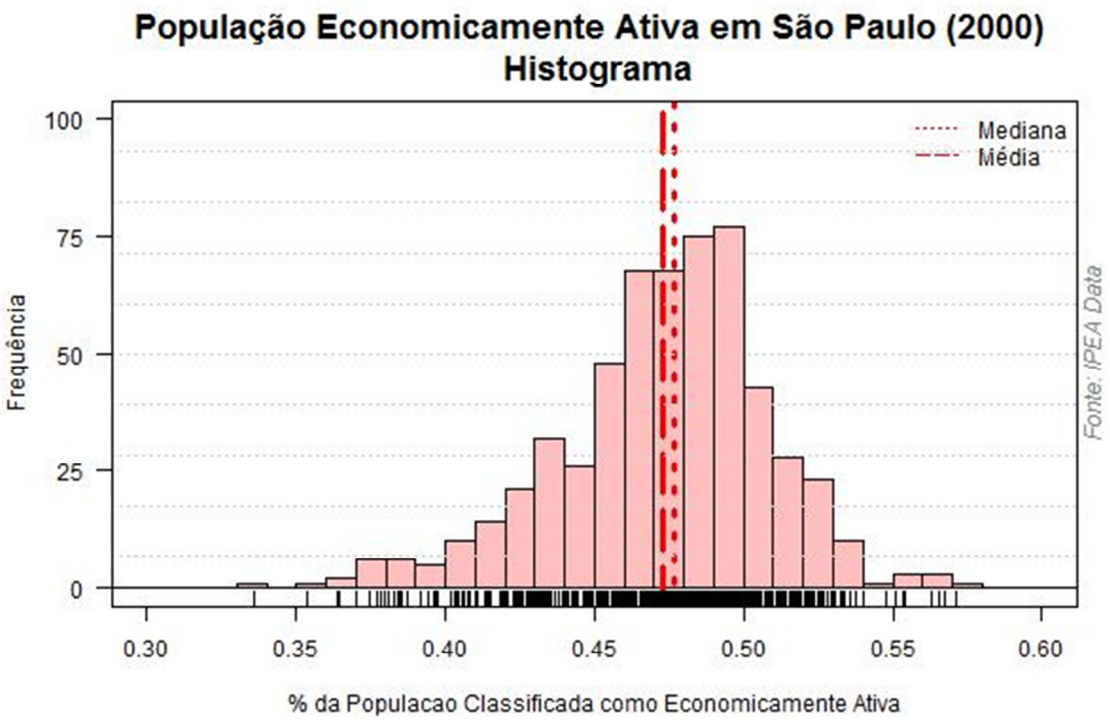
de Janeiro, chegando aos 35% da sua população disposta a trabalhar. No ano 2000, o Distrito Federal chegou a 50% de PEA.

1.2.3. São Paulo e o PEA

O gráfico 20, apresenta como a população de São Paulo se distribui, dentro de cada município, a partir da variável PEA

O gráfico abaixo é um histograma de frequência referente ao ano de 2000, com um pico em torno de 50%. Ou seja, 75 municípios possuem uma população de 50% de pessoas dispostas a trabalhar.

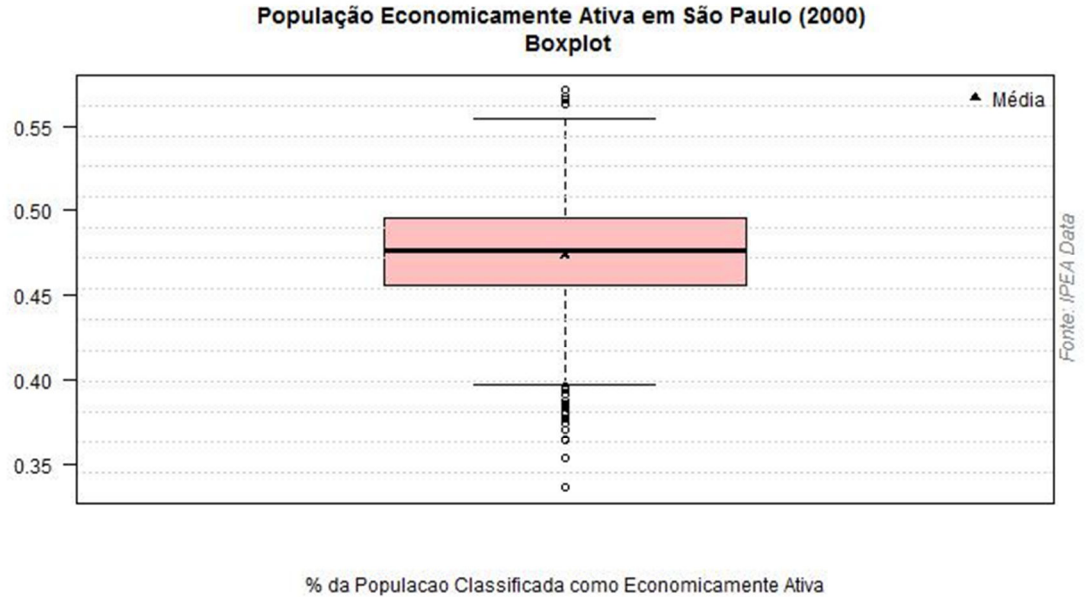
Gráfico 20



O formato do gráfico não se difere muito do gráfico nacional, e tem a média próxima da mediana. O *box plot* a seguir demonstra como a posição das observações encontradas em São Paulo e complementa a análise do histograma acima.

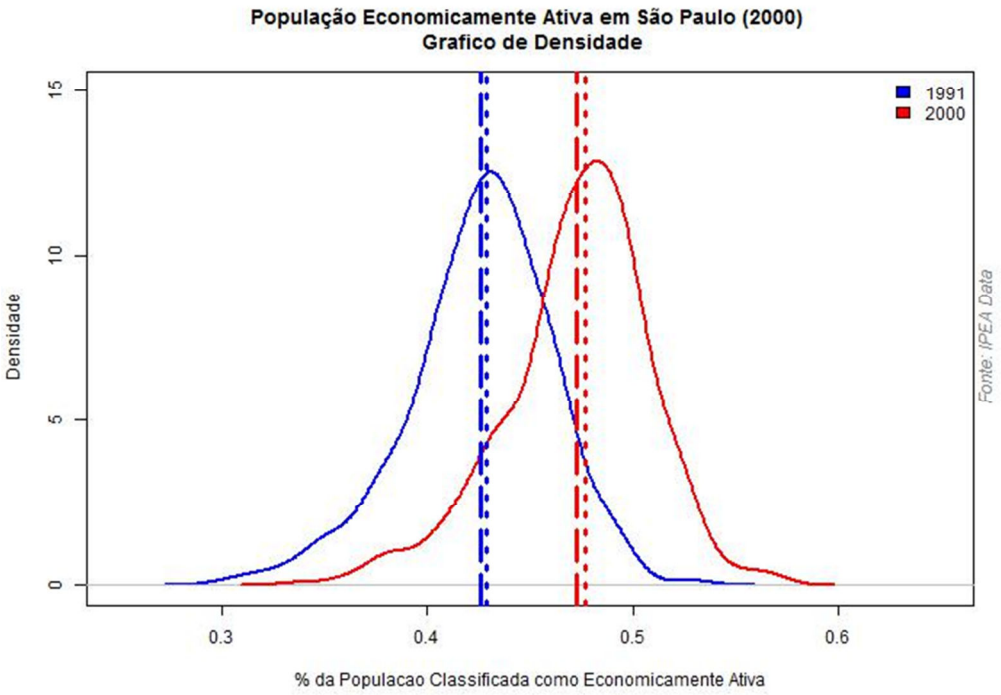
Observa-se no *box plot*, que os *outliers* estão abaixo de 40% e acima de 55% do PEA. Ou seja, um número não significativo de municípios teve uma população, em 2000, de menos de 40% da população disposta a trabalhar e outros tiveram uma população acima de 55% dispostas a trabalhar. A mediana e a média encontram-se próximos dos 47%, um pouco acima da média nacional. Dentre todos os estados, São Paulo, junto com o Rio de Janeiro, possui os dados menos dispersos.

Gráfico 21



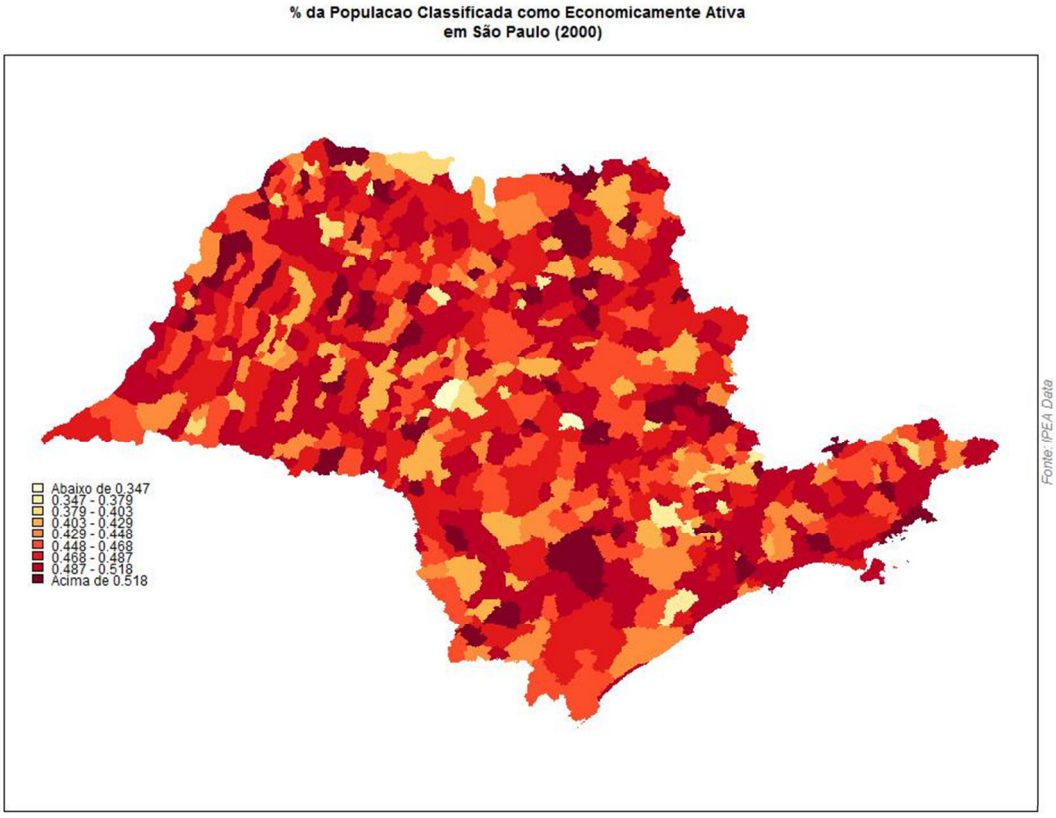
O gráfico 22 abaixo nos ajuda a entender a evolução do PEA dentro do estado de São Paulo, a partir de 1991 até o ano de 2000. O gráfico 22 apresenta a distribuição dos dados do PEA do estado de São Paulo dos anos 1991 e 2000. Verifica-se que São Paulo manteve a tendência nacional de ter em 2000 uma população economicamente ativa maior que em 1991. Tanto o gráfico foi deslocado para a direita assim como a média e a mediana.

Gráfico 22



O gráfico 23 é um mapa do estado de São Paulo, com os limites de cada cidade e seu respectivo índice PEA.

Gráfico 23



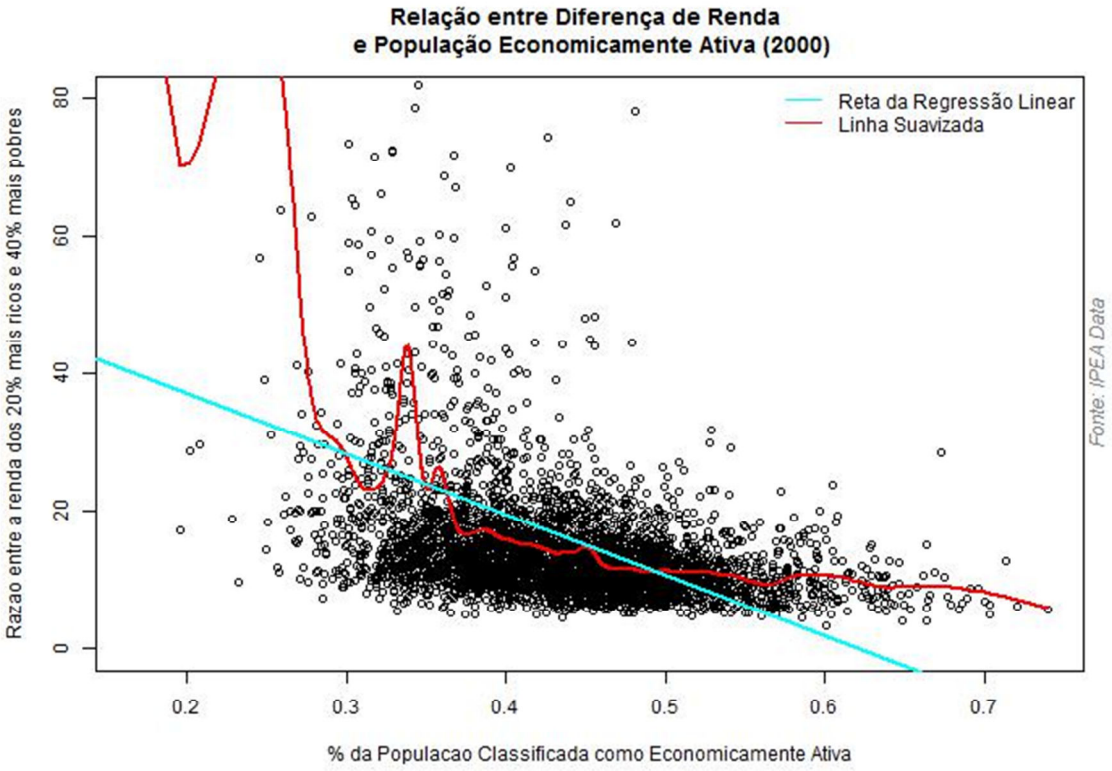
Quanto mais escuro o vermelho, mais pessoas dispostas a trabalhar naquele município. O inverso ocorre quanto mais claro for o vermelho, passando pelo amarelo antes de chegar ao branco.

O mapa corrobora a análise do *box plot* com a maior pare dos municípios com um PEA entre 40% e 50% e a presença de alguns *outliers* claros e escuros.

2. Análise das Variáveis Conjuntamente

2.1. Brasil

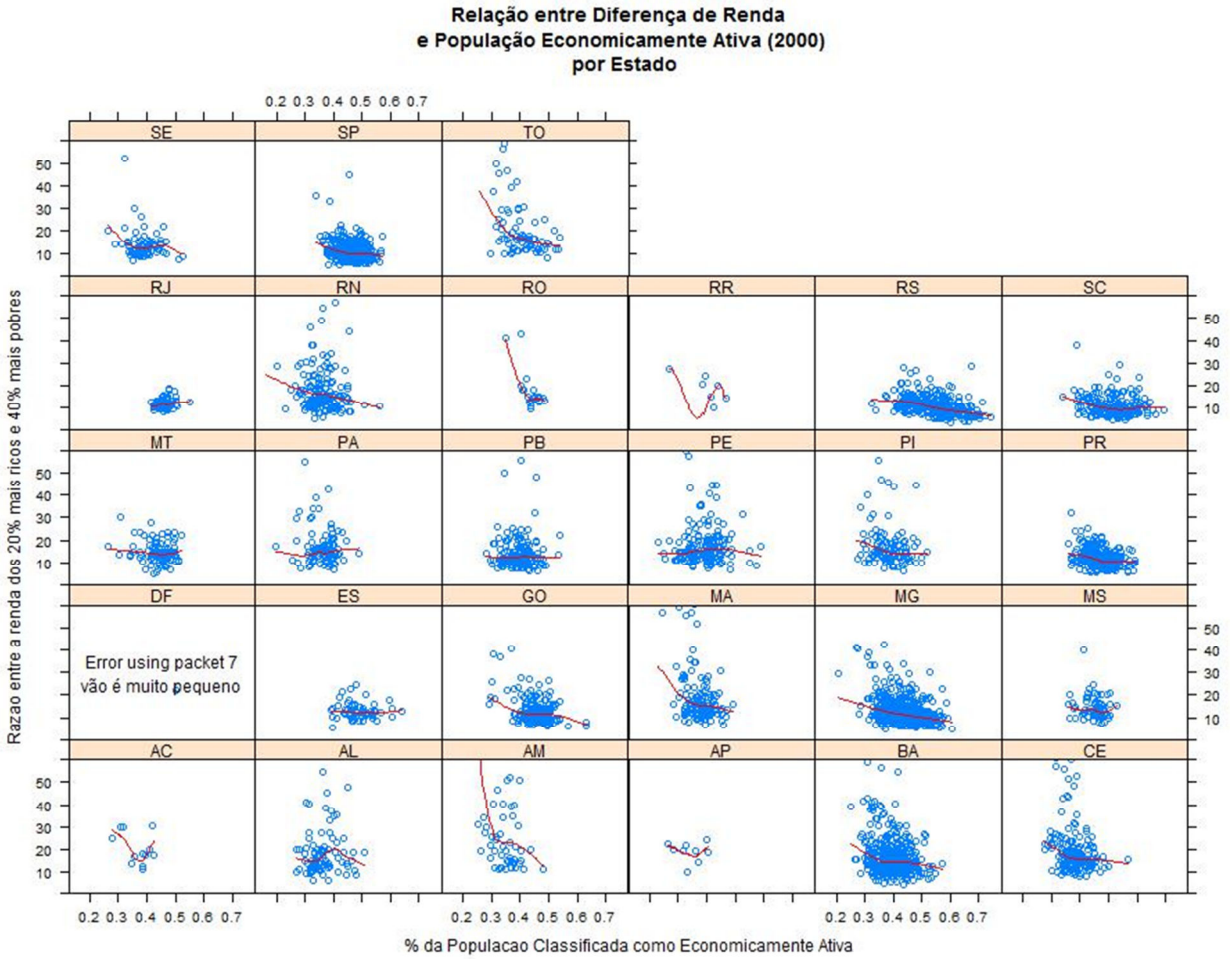
Gráfico 24



Os dados aparentam não ter relação, ou pequena relação, pois são bastante concentrados num valor baixo da razão de renda entre pobres e ricos. Ao realizar um teste de regressão linear, observamos um R2 de 0.02309, e um p-valor muitíssimo baixo, caracterizando assim que estas duas variáveis não possuem relação linear entre si, isto pode ser visto de forma visual no gráfico pela reta azul/verde, que representa a melhor aproximação que pode ser feita, que não atende ao comportamento dos dados. Talvez possamos ter alguma relação exponencial ou de outra natureza, como que quanto menor é a porcentagem da população que é economicamente ativa, maior a diferença nas rendas, mas os alunos ainda não sabem aplicar testes de outros tipos para verificar relações não-lineares.

2.2. Estados

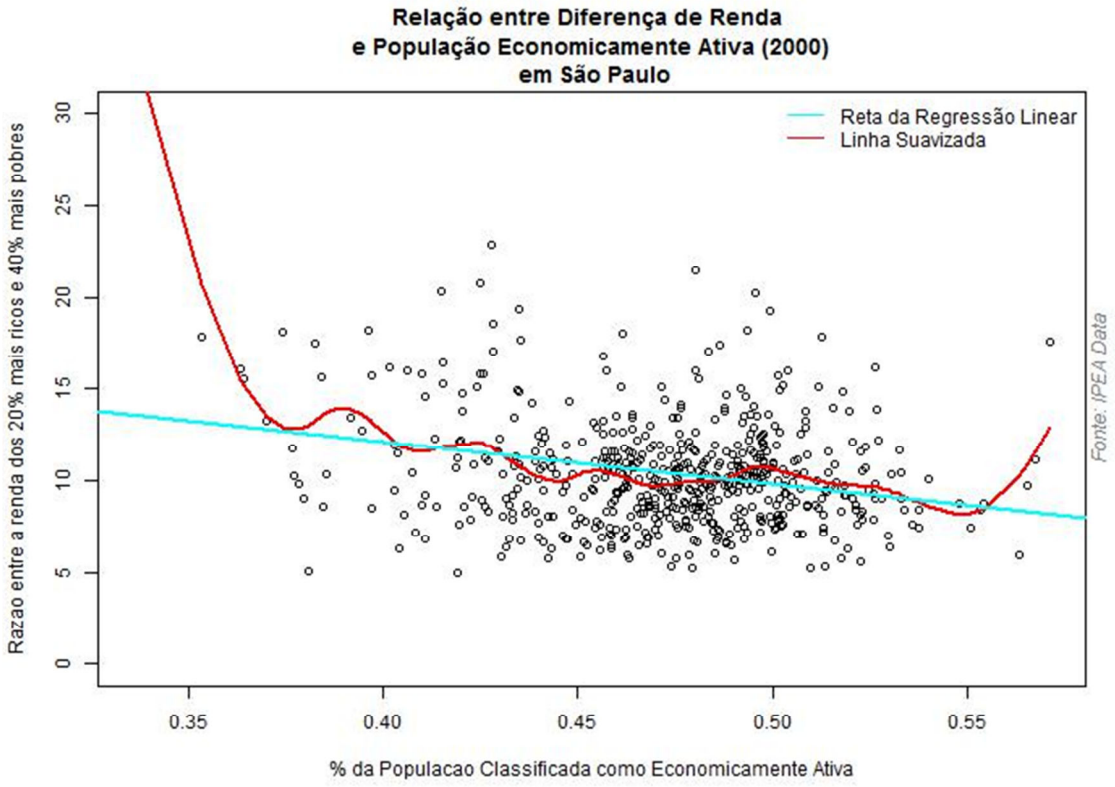
Gráfico 25



Podemos perceber que não há um padrão para todos os estados. Não são todos os estados que possuem variação na taxa da diferença da renda, como mostra o gráfico do Brasil, São Paulo, Rio de Janeiro, Rio Grande do Sul, Mato Grosso, Paraná, Mato Grosso do Sul e Amapá não possuem muita variação na vertical, ou seja, na diferença de renda. A grande maioria dos estados possui variação parecida no que diz respeito ao eixo horizontal, a porcentagem da população que é economicamente ativa. Os estados de Tocantins, Rio Grande do Norte, Paraná e Minas Gerais apresentam dispersão similar à dispersão do gráfico a nível de país. O estado do Rio de Janeiro apresenta um caso relevante de concentração maior do que a média em ambas as variáveis.

2.3. São Paulo

Gráfico 26



O estado de São Paulo apresenta uma dispersão mais uniforme do que o Brasil nesta relação de variáveis. Ao realizar o teste de Regressão Linear, temos um R2 de 0.05062 e um p-valor insignificante, mostrando que não há relação linear entre estas variáveis. Apesar das linhas da regressão linear e suavizada estarem próximas, e o gráfico de resíduos do método linear mostrar dispersão ao redor de zero, as variáveis não ter relação pois a “reta” que representa a relação não ter inclinação diferente de zero estatisticamente.

3. Conclusão

Na década de 90, viu-se que, felizmente, mais pessoas estavam dispostas a trabalhar. O número de pessoas ociosas e que não estavam dispostas a trabalhar era de 63% no inicio da década e esse número baixo para 57% no final da década.

Contudo, mais pessoas dispostas a trabalhar não se refletiu em uma melhora na distribuição de renda. A diferença entre os ricos e os pobres aumentou. No inicio da década, a diferença entre os 20% mais ricos e os 40% mais pobres era de aproximadamente 10, enquanto que o Brasil fechou a década com uma razão próxima a 15.

Como foi visto na análise conjunta, não se pode dizer que se o Brasil voltasse atrás e reduzisse o seu PEA, o problema da concentração de renda estaria resolvido. Já que a variável PEA, estatisticamente, não influencia a razão entre os mais ricos e pobres.

Apesar de não termos encontrado evidências de relação linear entre as variáveis, nossa análise visual indica que pode haver algum tipo de relação, possivelmente exponencial entre as variáveis, como que quanto menor é a porcentagem da população que é economicamente ativa, maior a diferença nas rendas. Para realizar estas análises precisaríamos conhecer outros modelos de testes de relações, e testá-los até encontrar algum adequado ou então definir que não há relação entre as variáveis.

É importante observar que não só os indícios visuais são importantes na análise dos dados, pois muitas vezes o gráfico pode aparentar alguma conclusão mas esta não é estatisticamente relevante, para isso devemos também utilizar testes estatísticos e outras análises computacionais que complementam a interpretação dos dados.