

Análise descritiva de dados

Vanessa F. Sehaber

17 de outubro de 2012

1 Mortalidade infantil até 5 anos e renda familiar per capita média no Brasil

O presente trabalho visa uma análise descritiva de dados de mortalidade infantil até 5 anos de idade e da renda familiar per capita média à nível de municípios do Brasil, dados estes obtidos do site www.ipeadata.gov.br, utilizando o *software* estatístico R.

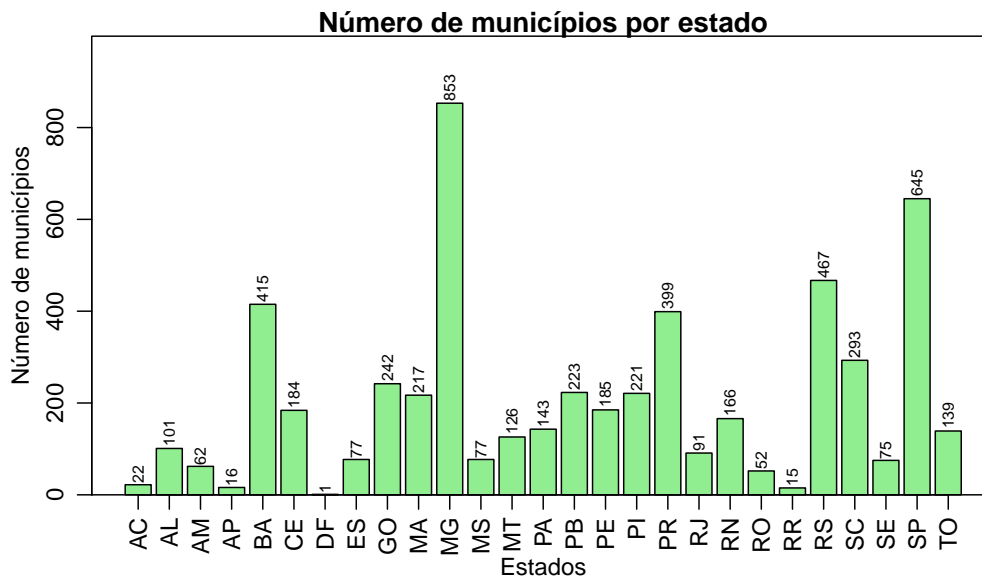
Após exclusão de valores que faltam nos bancos de dados, os dados de mortalidade infantil até 5 anos coletados nos municípios brasileiros têm 5507 observações e 5 variáveis. Os nomes dessas variáveis são: sigla do estado (*sig*), código do município (*cod*), nome do município (*mun*), taxa de mortalidade no ano de 1991 a cada mil nascidos (*x1*), taxa de mortalidade no ano de 2000 a cada mil nascidos (*x2*).

```
> head(mort.inf)
```

	sig	cod	mun	x1	x2
1	AC	1200013	Acrelândia	47.742	31.493
2	AC	1200054	Assis Brasil	65.415	30.702
3	AC	1200104	Brasiléia	77.233	34.133
4	AC	1200138	Bujari	40.267	31.493
5	AC	1200179	Capixaba	62.727	46.154
6	AC	1200203	Cruzeiro do Sul	48.890	36.261

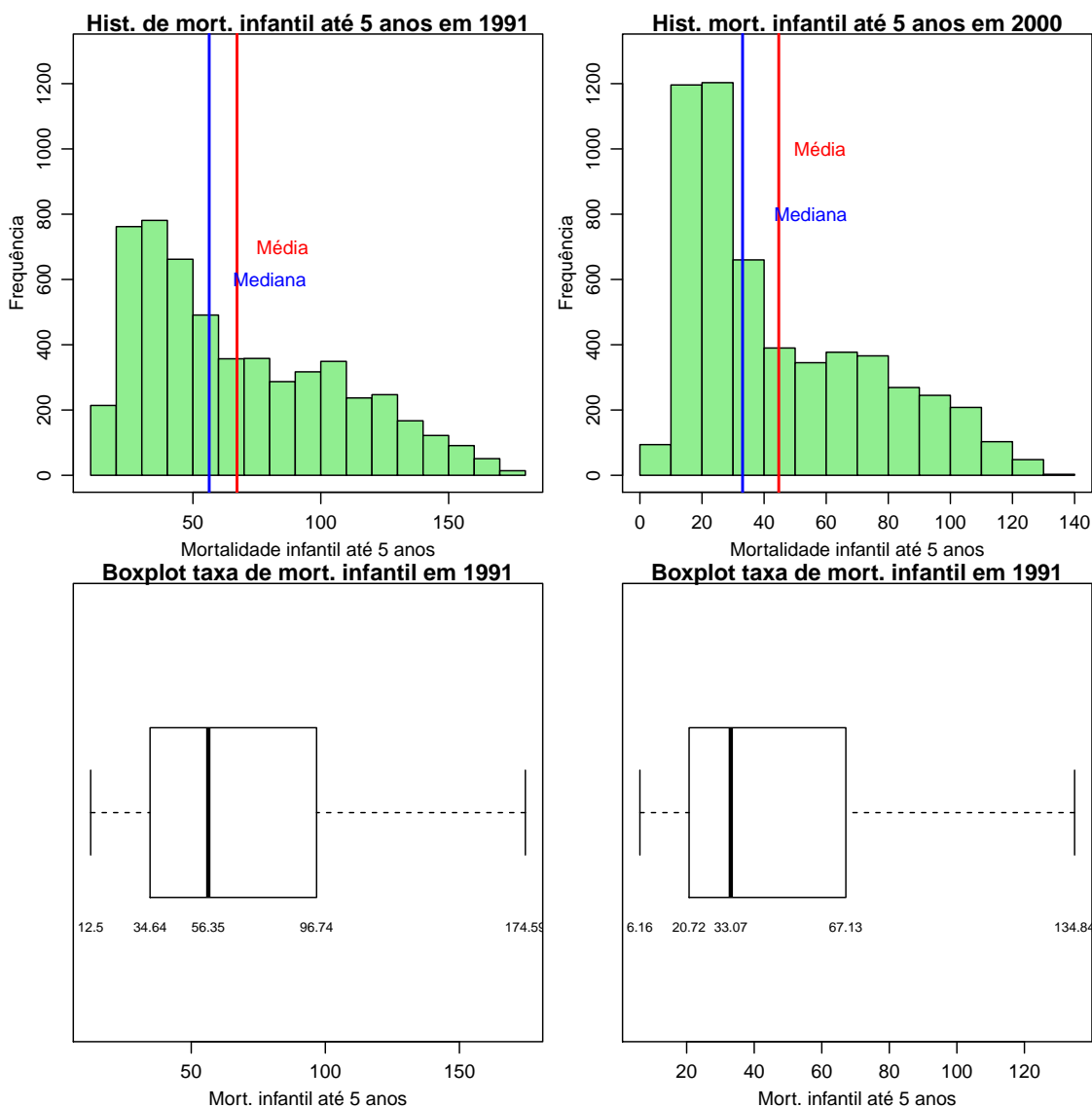
Em nível de Brasil, podemos tomar noção do comportamento e das informações que as variáveis disponibilizam por meio de alguns gráficos.

Primeiramente, vamos observar a quantidade de municípios por estado que estão sendo considerados nesse estudo.



Pela figura (1), podemos observar que os três estados que se destacam em relação à quantidade de municípios são os estados de Minas Gerais, São Paulo e Rio Grande do Sul e que os três estados com menos municípios são Distrito Federal, Roraima e Amapá.

As variáveis de interesse são x_1 e x_2 . Para avaliarmos as informações dessas variáveis, construímos dois gráficos: o histograma e o *boxplot*. O histograma é um gráfico composto por retângulos justapostos em que a base de cada um deles corresponde ao intervalo de classe e a sua altura corresponde à respectiva frequência, ou seja, a área de cada retângulo é proporcional à frequência da classe correspondente. O *boxplot* é uma simples representação em forma de diagrama dos cinco números sumários: mínimo, quartil inferior (primeiro quartil), mediana (segundo quartil), quartil superior (terceiro quartil) e máximo. Esses números são importantes para avaliarmos a dispersão dos dados.



Observando os histogramas e os *boxplots* da figura (1), percebe-se que a disposição das taxas de mortalidades infantis até 5 anos nos municípios são assimétricas em relação a média e que se concentram mais à esquerda do que à direita. Nestes gráficos, há duas medidas de posição as quais indicam a localização dos dados: a média e a mediana. A média é o valor que aponta para onde mais se concentram os dados de uma distribuição e a mediana indica o centro da distribuição dos dados, separando os mesmos em metade superior da metade inferior.

Dessas variáveis, vamos extrair algumas medidas estatísticas que são identificadas nos gráficos da

figura (1), sendo elas:

Tabela 1: Medidas estatísticas referente a variável x1.

Med. Estat.	Valores
Mín.	12.50
1 Qu.	34.64
Mediana	56.35
3 Qu.	96.74
Máx.	174.60
Média	67.23
Desv. Pad.	38.54

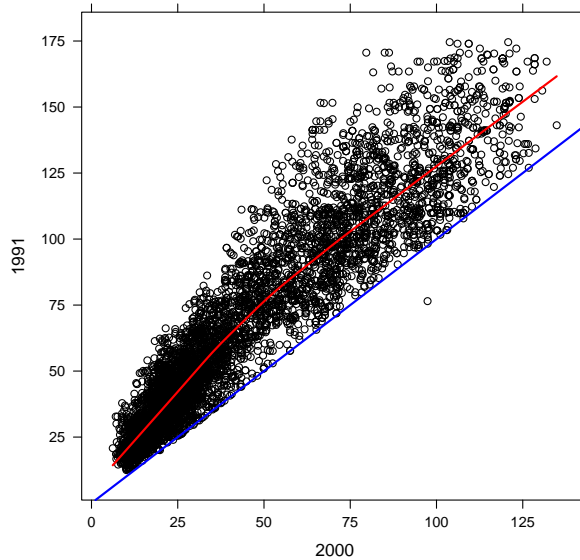
Tabela 2: Medidas estatísticas referente a variável x2.

Med. Estat.	Valores
Mín.	6.16
1 Qu.	20.72
Mediana	33.07
3 Qu.	67.13
Max.	134.80
Média	44.72
Desv. Pad.	29.72

O desvio padrão é uma medida que descreve a variação dos dados em relação à média. Assim, observa-se que os dados de mortalidade infantil até 5 anos em 1991 se desviam mais da sua respectiva média do que o dados do ano de 2000, pois $\sigma_{1991} = 38,54 > \sigma_{2000} = 29,72$. Máximo e mínimo são utilizados para encontrar a amplitude dos dados. Logo, a amplitude dos dados de 1991 é $a_{1991} = 162,1$ e dos dados de 2000 é $a_{2000} = 128,64$. Quartis são medidas para sumarizar os dados, dividindo os mesmos em quatro partes iguais, sendo eles 1º quartil, 2º quartil (ou mediana) e o 3º quartil. Comparando as duas tabelas acima, verifica-se que os dados estão mais concentrados à esquerda, e que no ano de 2000 as medidas estatísticas diminuíram, ou seja, houve alguma intervenção durante 1991 e 2000 que ajudou a diminuir as taxas de mortalidade infantil até 5 anos nos municípios brasileiros.

Para sabermos como as taxas de mortalidade infantil até 5 anos nesses municípios se dispuseram em 1991 e 2000, vamos fazer um gráfico de dispersão entre essas variáveis e traçarmos algumas linhas de tendência a fim de analisarmos a relação entre tais variáveis.

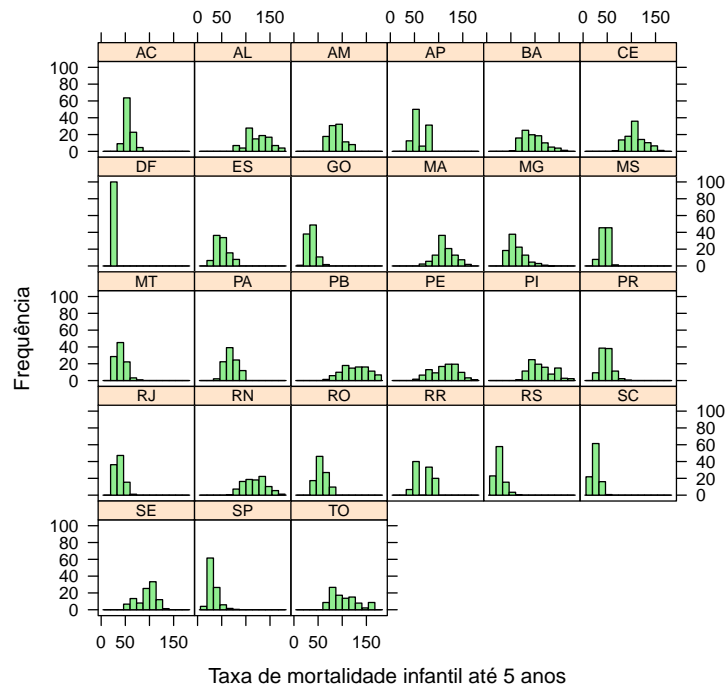
Dispersão dos dados de mortalidade anos 1991 x 2000



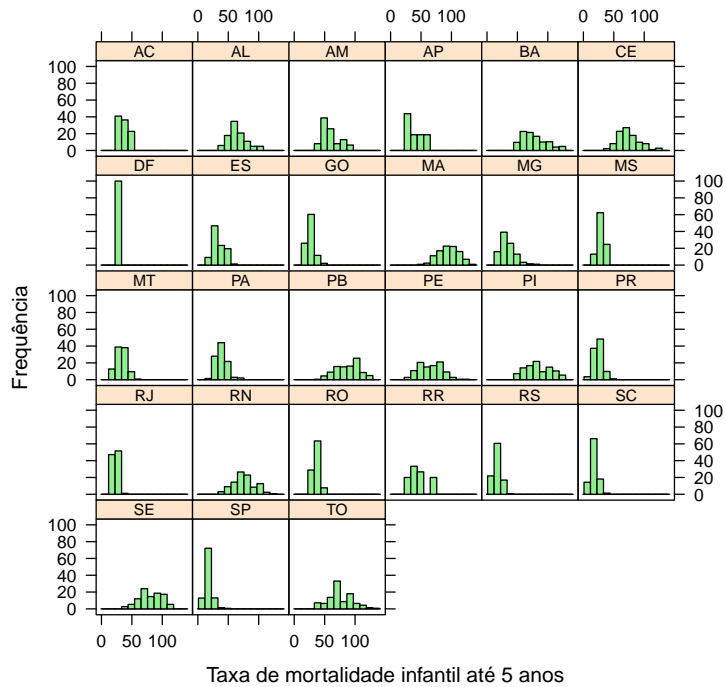
A figura (1) mostra que as taxas de mortalidade infantil até 5 anos possuem relação (a linha de tendência vermelha sugere uma relação não-linear) e que a dispersão maior se verifica no ano de 1991 (amplitude do eixo). Ainda, apenas uma cidade teve a taxa de mortalidade infantil até 5 anos maior no ano de 2000 (ponto abaixo da linha reta) enquanto que as demais cidades tiveram tal taxa abaixo.

A seguir serão apresentados dois gráficos pelos quais podemos comparar como se comportaram as taxas de mortalidade infantil até 5 anos do ano de 1991 para o ano de 2000 em cada estado.

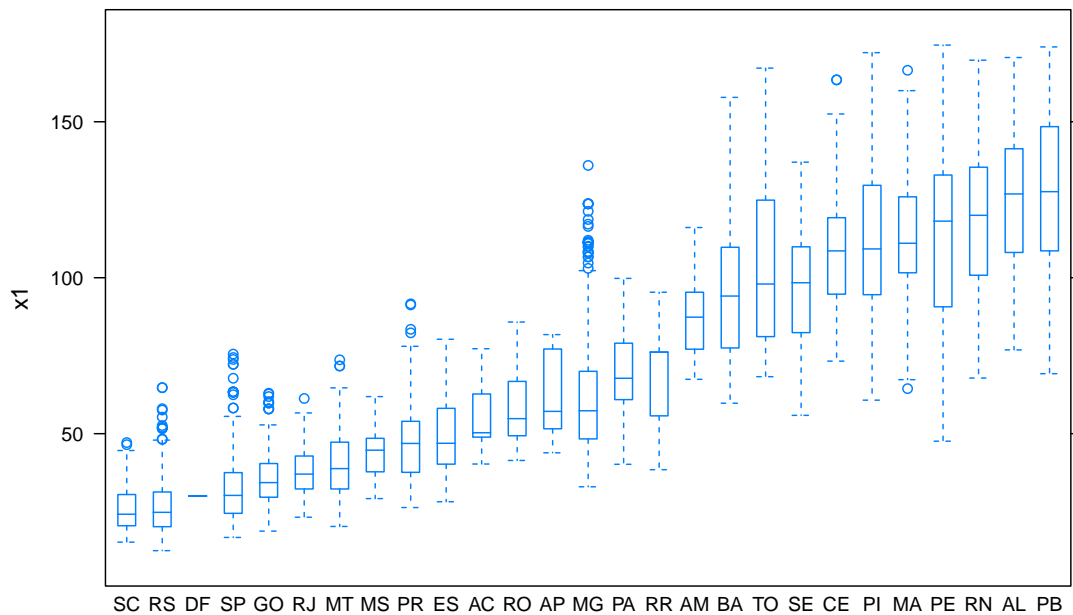
Hist. da mort. infantil até 5 anos em 1991 por estado

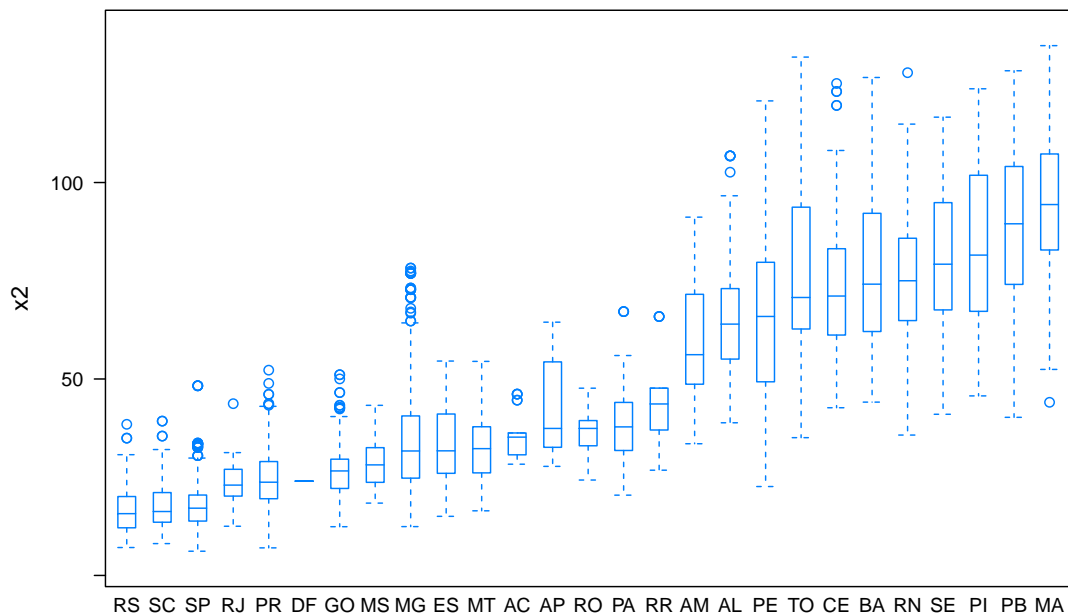


Hist. da mort. infantil até 5 anos em 2000 por estado



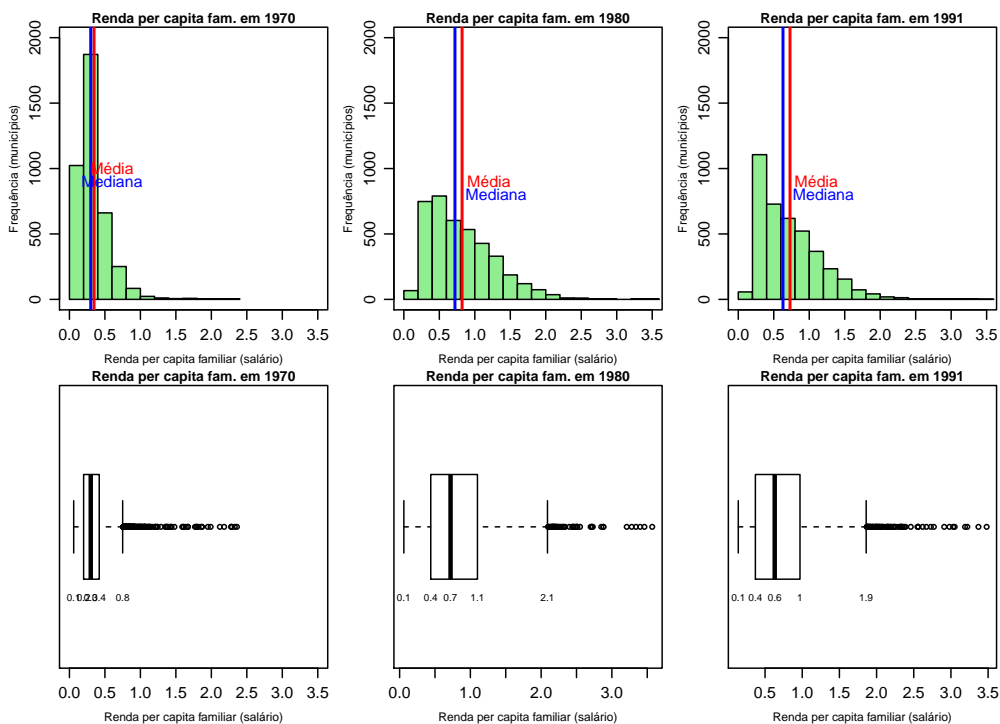
De modo geral, podemos dizer que a maioria das taxas de mortalidade infantil até 5 anos dos estados brasileiros diminuiu, em especial as taxas de Pernambuco e Alagoas. Porém, as taxas nos estados da Bahia, Maranhão e Sergipe diminuíram pouco dando a falsa impressão de que as mesmas aumentaram em relação a dos demais estados. Essas comparações ficam mais evidentes quando observamos os gráficos *boxplot* a seguir:





Observa-se que o gráfico da figura (1) possui mais pontos em destaque em relação ao gráfico da figura (1), o que pode indicar que, no período compreendido entre 1991 e 2000, algumas cidades não conseguiram diminuir suas taxas de mortalidade infantil até 5 anos, ou a diminuição foi muito pequena.

Vamos apresentar agora como a renda per capita familiar média tem se comportado no período de 1970 a 1991 (o site do Ipeadata não disponibilizou dados referentes ao ano de 2000). Excluindo os valores não informados, foram considerados 3951 municípios para realização do gráfico a seguir.

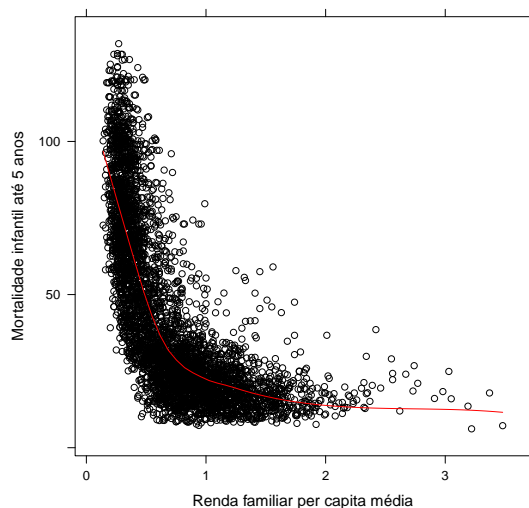


Os gráficos apresentados na figura (1) mostram que a renda familiar per capita média nos municípios

brasileiros aumentou entre 1970 e 1991. Em 1970, a renda familiar per capita média se concentrava mais entre 0 – 0,5 e poucas cidades possuíam essa renda acima de 1, o que pode ser interpretado como pontos discrepantes no respectivo *boxplot*. Já em 1980, a renda familiar per capita média se dispersou mais à direita, e aumentou a quantidade de municípios que possuíam a renda familiar per capita média acima de 1. Porém em 1991, a renda familiar per capita média dos municípios acaba diminuindo um pouco e sugere um crescimento dos pontos discrepantes, que são as cidades com renda familiar per capita média mais elevada em relação à grande maioria dos municípios brasileiros.

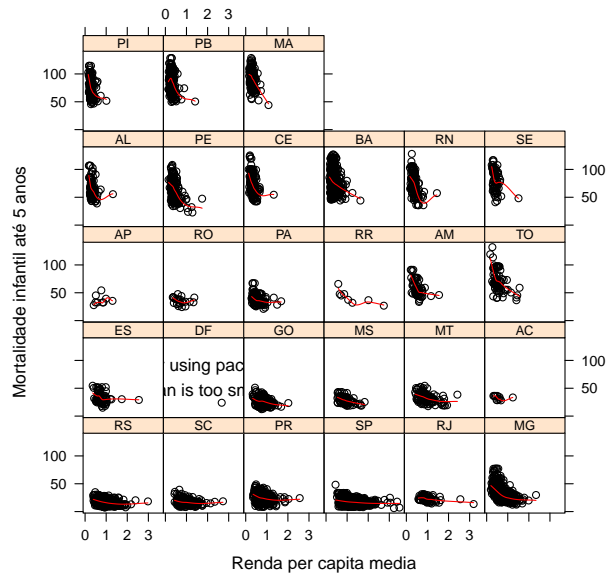
Considerando os dados municipais de mortalidade infantil até 5 anos de idade e de renda familiar per capita média do ano de 1991 à nível de Brasil, vamos fazer um gráfico de dispersão para verificarmos se há alguma relação entre tais dados. Para realizar as análises a seguir, utilizou-se 4491 municípios, em 1991, devido aos valores não informados no banco de dados disponibilizado.

Gráfico taxa mortalidade x renda per capita média (1991)



O gráfico da figura (1) mostra que as duas variáveis consideradas se relacionam não-linearmente, ou seja, que as maiores taxas de mortalidade infantil até 5 anos de idade tem relação com a renda familiar per capita média baixa e, conforme a renda familiar per capita média aumenta as taxas de mortalidade infantil até 5 anos de idade diminuem.

Taxa mortal. x renda per capita média por estados (1991)



À nível de estados, podemos verificar que os estados de Alagoas, Amazônia, Bahia, Ceará, Maranhão, Piauí, Pernambuco, Paraíba, Rondônia, Tocantins e Sergipe são os estados onde as maiores taxas de mortalidade infantil até 5 anos de idade tem relação com a renda familiar per capita média baixa. E que os demais estados, onde a renda familiar per capita média é maior, as taxas de mortalidade infantil até os 5 anos de idade são menores, exceto os estados de Minas Gerais e de Roraima, onde verificam-se os dois tipos de situações.