

Visualização de Dados

Autor: Luciano Moraes Da Luz Brum.
Professor: Dr. Sandro Camargo.

Introdução

- Base de dados: Dados de produtores de leite do município de Derrubadas – RS dos anos de 2014, 2015 e 2016.
- Os dados desta planilha eram numéricos, com 167 amostras e 37 atributos e alguns parâmetros destes.

Introdução

Município: Derrubadas - RS																													
		Produtor		ID-UF	S1	S2	S3	S4	S5	S6	E1	E2	E3	E4	E5	E6	E7	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	
1	1	Produtor 1		2014	0,62	0,50	0,67	0,60	1,00	1,00	1,00	0,34	0,00	0,36	0,10	0,60	0,51	1,00	1,00	1,00	0,00	1,00	0,50	0,00	0,67	1,00	0,50	1,00	1,00
1	1	Produtor 1		2015	0,54	0,82	0,67	0,60	0,00	0,00	1,00	0,49	0,00	0,36	0,10	0,80	0,29	1,00	1,00	1,00	0,00	1,00	0,50	0,00	0,67	1,00	0,50	0,00	1,00
1	1	Produtor 1		2016	0,54	0,82	0,67	0,60	0,00	0,00	1,00	0,49	0,00	0,36	0,10	0,80	0,29	1,00	1,00	1,00	0,00	1,00	0,50	0,00	0,67	1,00	0,50	0,00	1,00
17	2	Produtor 2		2014	0,64	0,80	0,67	0,60	1,00	0,00	1,00	0,37	0,00	0,52	0,14	0,80	0,29	1,00	1,00	1,00	0,00	0,00	0,50	0,00	1,00	0,90	0,50	1,00	1,00
17	2	Produtor 2		2015	0,60	0,50	0,67	0,60	1,00	0,00	1,00	0,35	0,00	0,40	0,14	0,85	0,26	0,93	1,00	1,00	0,00	0,00	0,50	0,00	1,00	0,90	0,50	1,00	1,00
17	2	Produtor 2		2016	0,60	0,50	0,67	0,60	1,00	0,00	1,00	0,35	0,00	0,40	0,14	0,85	0,26	0,93	1,00	1,00	0,00	0,00	0,50	0,00	1,00	0,90	0,50	1,00	1,00
17	3	Produtor 3		2014	0,45	0,25	0,67	0,60	1,00	1,00	1,00	0,18	0,00	0,31	0,18	0,80	1,00	0,10	1,00	1,00	0,00	1,00	0,50	0,00	0,67	1,00	0,50	0,00	0,00
19	4	Produtor 4		2015	0,63	0,83	0,67	0,60	1,00	0,00	1,00	0,52	0,00	0,18	0,10	0,85	0,50	1,00	1,00	1,00	0,00	1,00	1,00	0,50	0,67	0,90	0,50	1,00	1,00
19	4	Produtor 4		2016	0,60	0,80	0,67	0,60	1,00	0,00	1,00	0,31	0,00	0,18	0,10	0,85	0,00	1,00	1,00	1,00	0,00	1,00	1,00	0,50	0,67	0,90	0,50	1,00	1,00
19	4	Produtor 4		2014	0,59	0,50	0,67	0,60	1,00	0,00	1,00	0,52	0,00	0,18	0,18	0,60	0,50	1,00	1,00	1,00	0,00	1,00	1,00	0,50	0,47	0,90	0,50	0,50	1,00
22	5	Produtor 5		2015	0,51	0,00	0,67	0,60	1,00	0,00	1,00	0,69	0,00	0,22	0,07	0,60	0,17	0,83	0,60	1,00	0,00	0,00	0,50	0,50	0,80	0,65	0,50	1,00	0,00
22	5	Produtor 5		2016	0,51	0,00	0,67	0,60	1,00	0,00	1,00	0,71	0,00	0,18	0,07	0,60	0,19	0,84	0,60	1,00	0,00	0,00	0,50	0,50	0,80	0,65	0,50	1,00	0,00
22	5	Produtor 5		2014	0,45	0,00	0,67	0,60	1,00	1,00	1,00	0,12	0,00	0,22	0,07	0,60	1,00	0,00	0,60	1,00	0,00	0,00	0,50	0,50	0,80	0,65	0,50	1,00	0,00
21	6	Produtor 6		2016	0,60	0,25	0,67	0,60	1,00	0,00	0,00	0,60	0,00	0,09	0,17	0,85	0,29	1,00	1,00	1,00	0,00	0,00	1,00	0,50	0,73	0,90	0,50	1,00	1,00
21	6	Produtor 6		2015	0,59	0,25	0,67	0,60	1,00	0,00	0,00	0,55	0,00	0,09	0,09	0,85	0,21	1,00	1,00	1,00	0,00	0,00	1,00	0,50	0,73	0,90	0,50	1,00	1,00
21	6	Produtor 6		2014	0,55	0,25	0,67	0,60	1,00	0,00	0,00	0,41	0,00	0,09	0,09	0,85	0,18	0,81	1,00	1,00	0,00	0,00	0,50	0,50	0,40	0,90	0,50	1,00	1,00
1	7	Produtor 7		2015	0,55	0,00	0,33	0,90	0,00	0,00	1,00	0,28	1,00	0,60	0,18	0,80	0,30	1,00	1,00	0,00	0,00	1,00	1,00	0,50	0,53	0,90	1,00	1,00	1,00
1	7	Produtor 7		2016	0,55	0,00	0,33	0,90	0,00	0,00	1,00	0,28	1,00	0,60	0,18	0,80	0,30	1,00	1,00	0,00	0,00	1,00	1,00	0,50	0,53	0,90	1,00	1,00	1,00

Figura 1: planilha com dados do município de Derrubadas – RS. Fonte: Elaborada pelo autor, 2017

Introdução

2			RENDA FAMILIAR PER CAPITA			SATISFAÇÃO	PRODUÇÃO	PRODUÇÃO	VALOR DE PREÇO	DE SEMEADURA	QUANTIDADE DE ATIVIDADE	PARTICIPAÇÃO DO LEITE NA RENDA			INDEPENDÊNCIA FINANCEIRA			
	Nome do produtor	Ano	Número de Membros da Família	Receita anual da família (R\$)	Renda familiar per capita (mensal)	Felicidade	Produção total de leite (l/dia)	Produção por vaca (l/dia)	Qtidade de vacas (secas + lactação)	Valor total dos dos rebanhos (R\$)	Quantidade de atividades (mín. 1 e máx. 8)	Receita com o Leite (R\$/ano)	Receita total da família (dentro + fora) em R\$	Renda do Leite (%)	Dívida anual	Receita Dentro (R\$)	Receita Fora (R\$)	Grau de endividamento (%)
3																		
4	Produtor 39	2016	4	105.810	2.204	8	225	15	16	48.720	4	54.000	105.810	51	4.200	95.250	10.560	4%
6	Produtor 43	2016	3	100.500	2.792	8	68	9	8	87.400	3	19.200	100.500	19	3.010	82.800	17.700	3%
10	Produtor 18	2016	4	61.650	1.284	8	80	8	15	28.820	5	18.360	61.650	30	0	50.300	11.350	0%
14	Produtor 40	2016	2	49.410	2.059	8	70	5	14	34.000	5	15.300	49.410	31	0	37.970	11.440	0%
15	Produtor 26	2016	3	43.460	1.207	6	32	8	6	17.000	4	12.000	43.460	28	200	18.620	24.840	0%
19	Produtor 9	2016	5	39.172	653	8	27	5	5	17.180	5	0	39.172	0	0	12.400	26.772	0%
20	Produtor 22	2016	3	38.580	1.072	8	35	5	14	45.100	4	9.600	38.580	25	3.000	28.980	9.600	8%
21	Produtor 27	2016	4	37.944	791	8	45	7	9	23.600	5	10.980	37.944	29	0	34.980	2.964	0%
22	Produtor 35	2016	4	37.293	777	8	91	13	7	15.500	5	22.204	37.293	60	0	34.804	2.489	0%
25	Produtor 24	2016	4	35.804	746	8	38	8	10	21.900	4	9.150	35.804	26	0	16.440	19.364	0%
27	Produtor 12	2016	3	35.740	993	8	45	8	7	16.110	5	9.600	35.740	27	1.000	14.620	21.120	3%
35	Produtor 13	2016	3	31.040	862	8	48	5	9	16.250	5	13.440	31.040	43	500	25.840	5.200	2%
39	Produtor 29	2016	4	30.405	633	8	23	5	5	12.070	5	5.685	30.405	19	0	30.405	0	
41	Produtor 42	2016	3	30.235	840	6	23	5	5	13.140	4	5.600	30.235	19	0	25.235	5.000	0%
42	Produtor 20	2016	4	29.890	623	6	0	0	7	13.400	5	0	29.890	0	1.152	7.920	21.970	4%
44	Produtor 2	2016	5	29.560	493	8	49	5	10	21.260	5	7.680	29.560	26	1.000	18.380	11.180	3%
45	Produtor 25	2016	4	29.400	613	6	40	6	7	22.450	4	7.680	29.400	26	500	8.280	21.120	2%
49	Produtor 59	2016	4	28.760	599	8	45	5	10	27.350	4	14.040	28.760	49	0	28.760	0	

Figura 2: planilha com dados do município de Derrubadas – RS. Fonte: Elaborada pelo autor, 2017

Objetivo

- Aplicar técnicas de visualização de dados, explorando algumas das diferentes possibilidades em gráfico 2D:
 - Posição;
 - Cor;
 - Forma;
 - Intensidade de cor ou brilho;
 - Rotação;
 - Tamanho;
 - Textura;
 - Movimento;

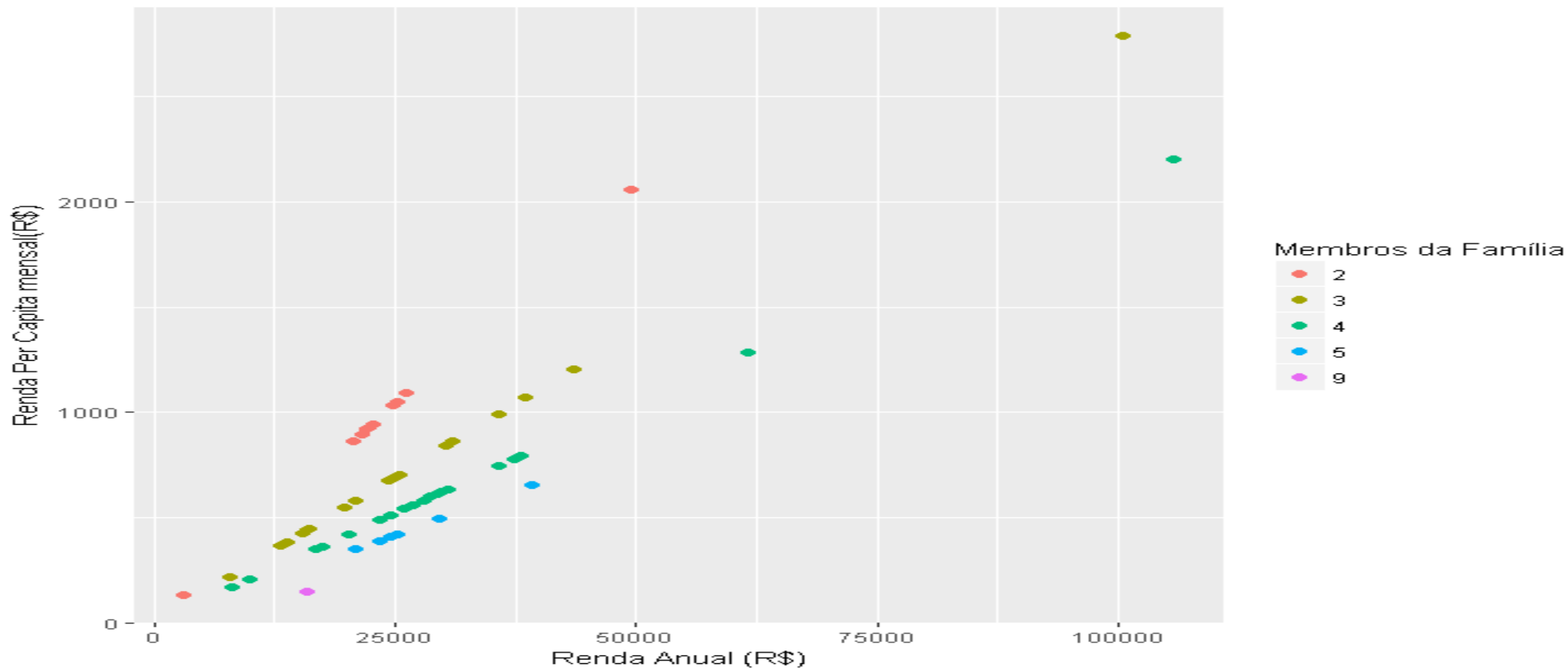
Passo a Passo

- Foi utilizada a ferramenta R para o processo de limpeza, categorização e visualização dos dados.
- Das instruções utilizadas no R:

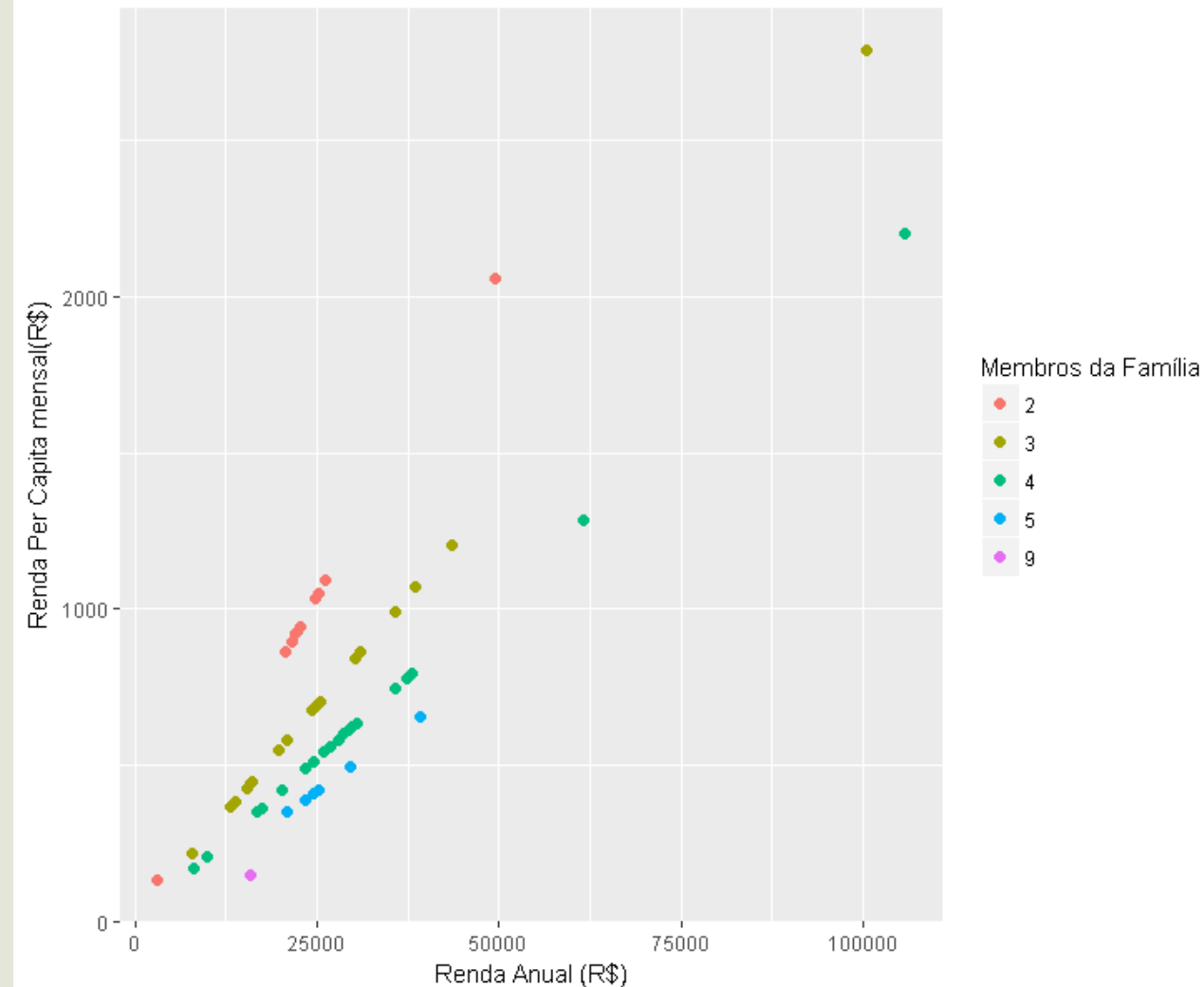
```
library(ggplot2)
n3 <- read.csv('leite2016.csv',sep=";",stringsAsFactors=FALSE)
for(i in 1:nrow(n3)){
  for(j in 3:ncol(n3)-1){
    n3[i,j] = as.numeric(n3[i,j])} }
```

```
for(i in 1:nrow(n3)){
  if(n3[i,ncol(n3)]==""){
    n3[i,ncol(n3)] = as.numeric(gsub("", "0", n3[i,ncol(n3)])) }
  else{
    n3[i,ncol(n3)] = as.numeric(gsub("[%]", "", n3[i,ncol(n3)]))} }
for(i in 1:nrow(n3)){
  n3[i,ncol(n3)]<-as.numeric(n3[i,ncol(n3)]/(100))
}
```

Passo a Passo

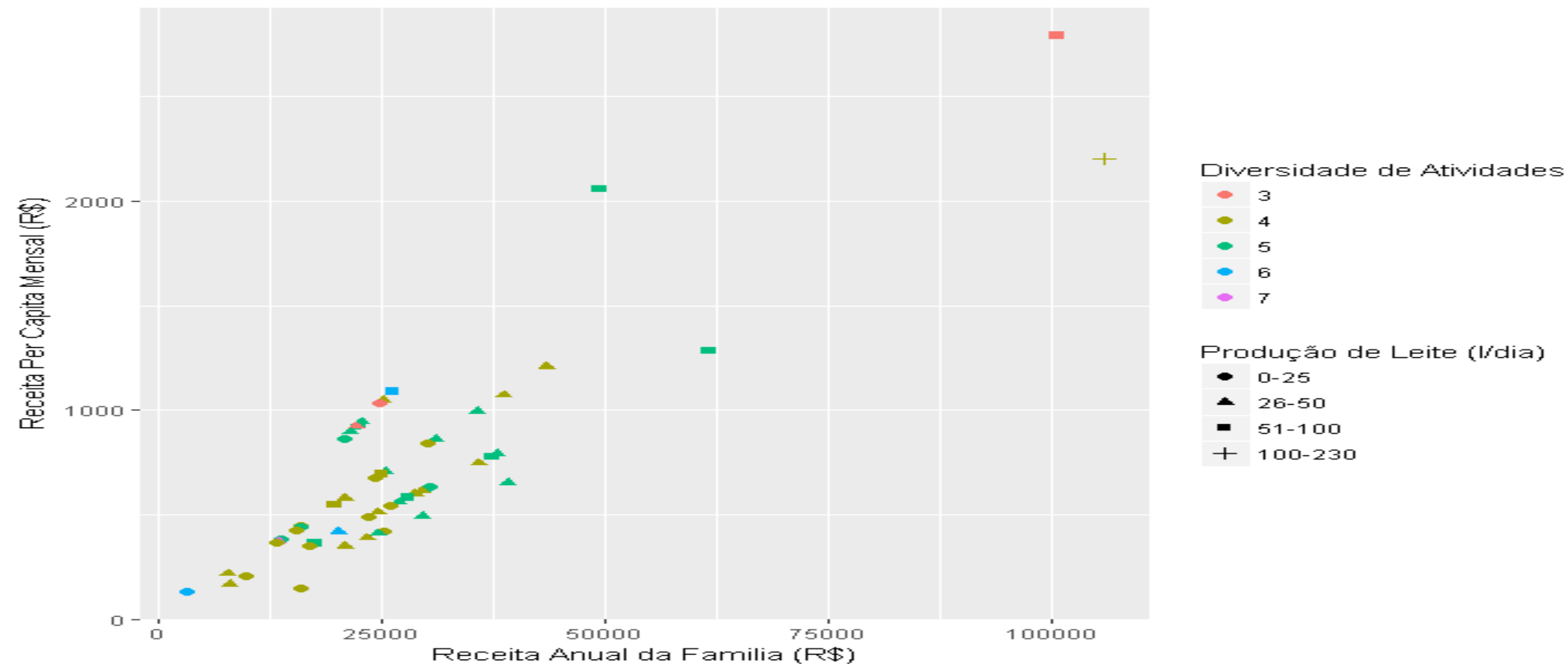


Passo a Passo

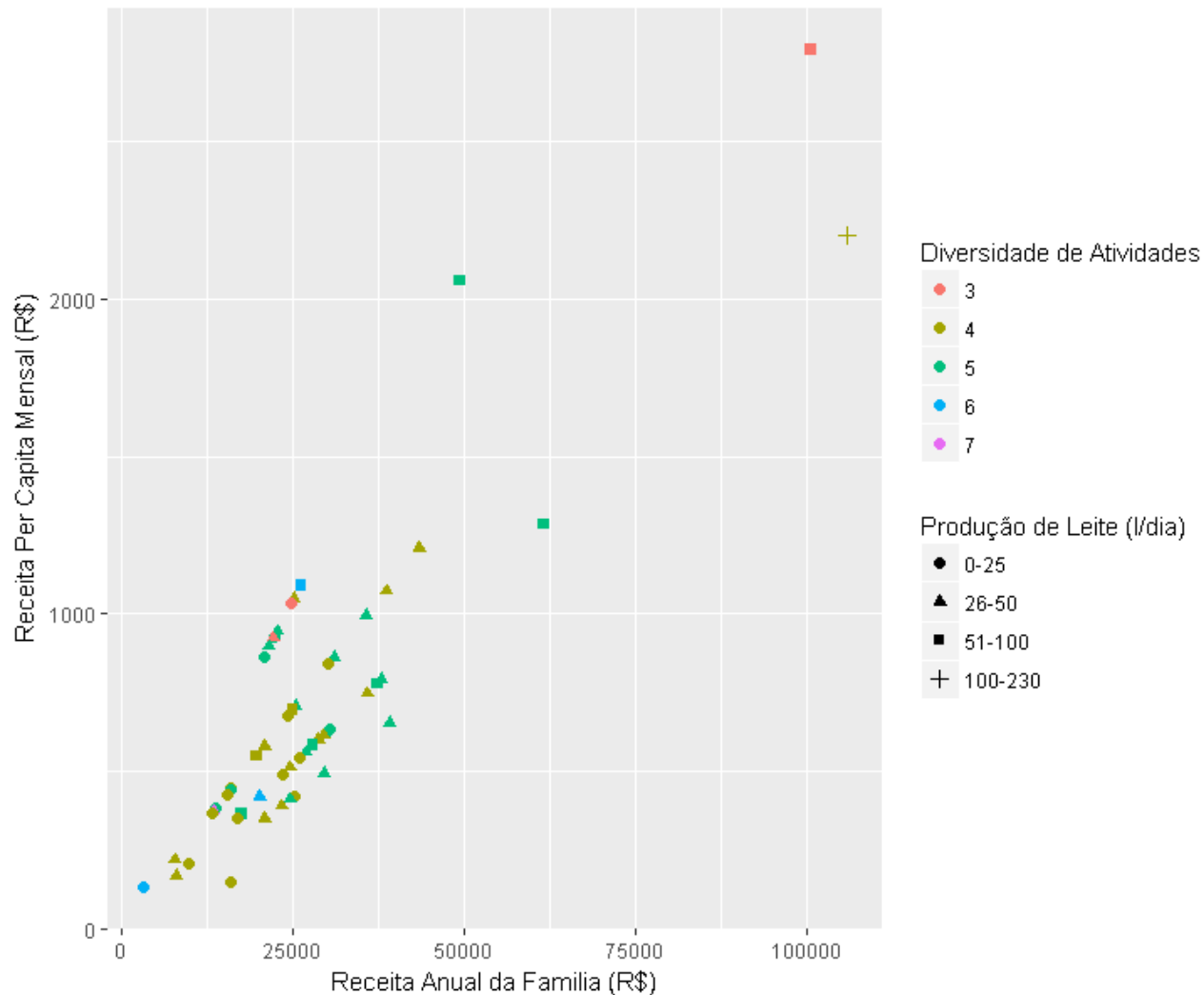


```
ggplot(n3,  
  aes(n3[,4],n3[,5],col=as.factor(n3[,3])))+  
  geom_point(size=2)+  
  labs(colour = "Membros da Família") +  
  scale_x_continuous(name="Renda Anual (R$)") +  
  scale_y_continuous(name="Renda Per Capita mensal(R$)")
```


Passo a Passo



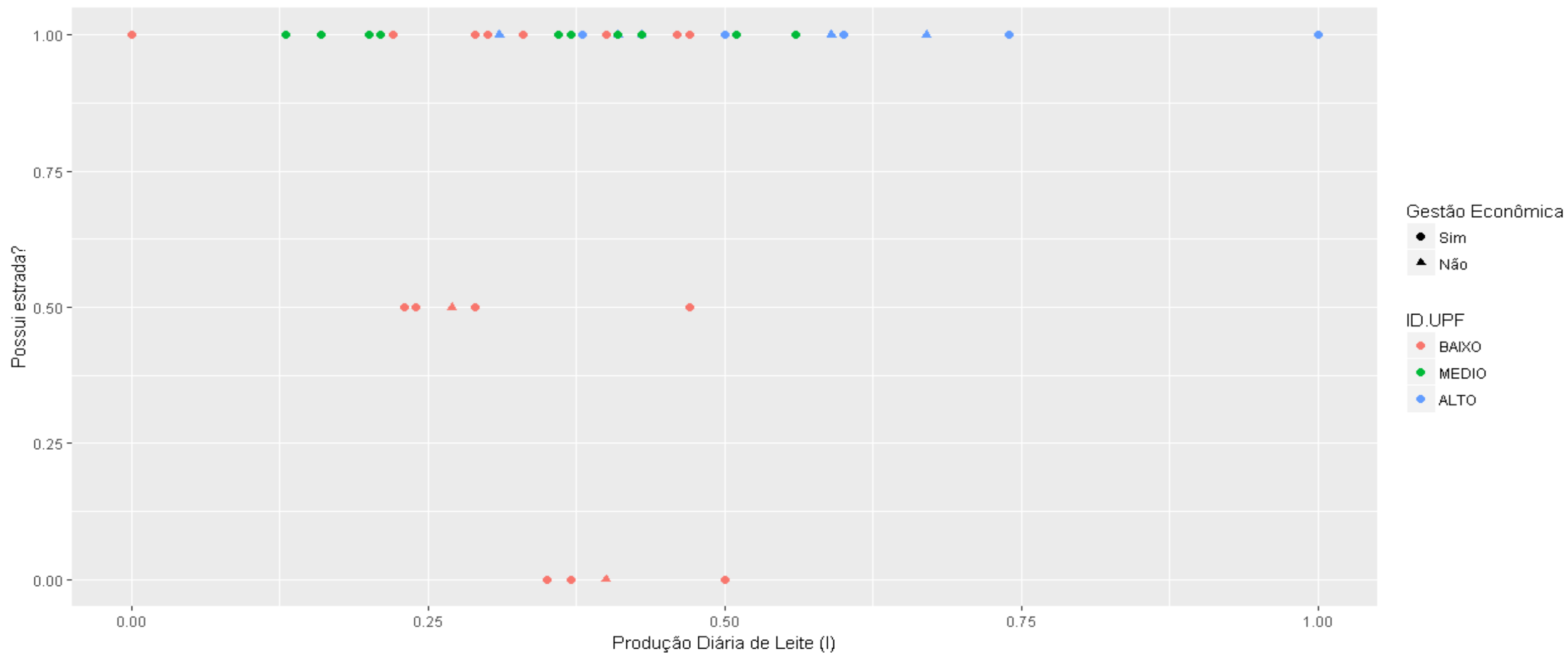
Passo a Passo



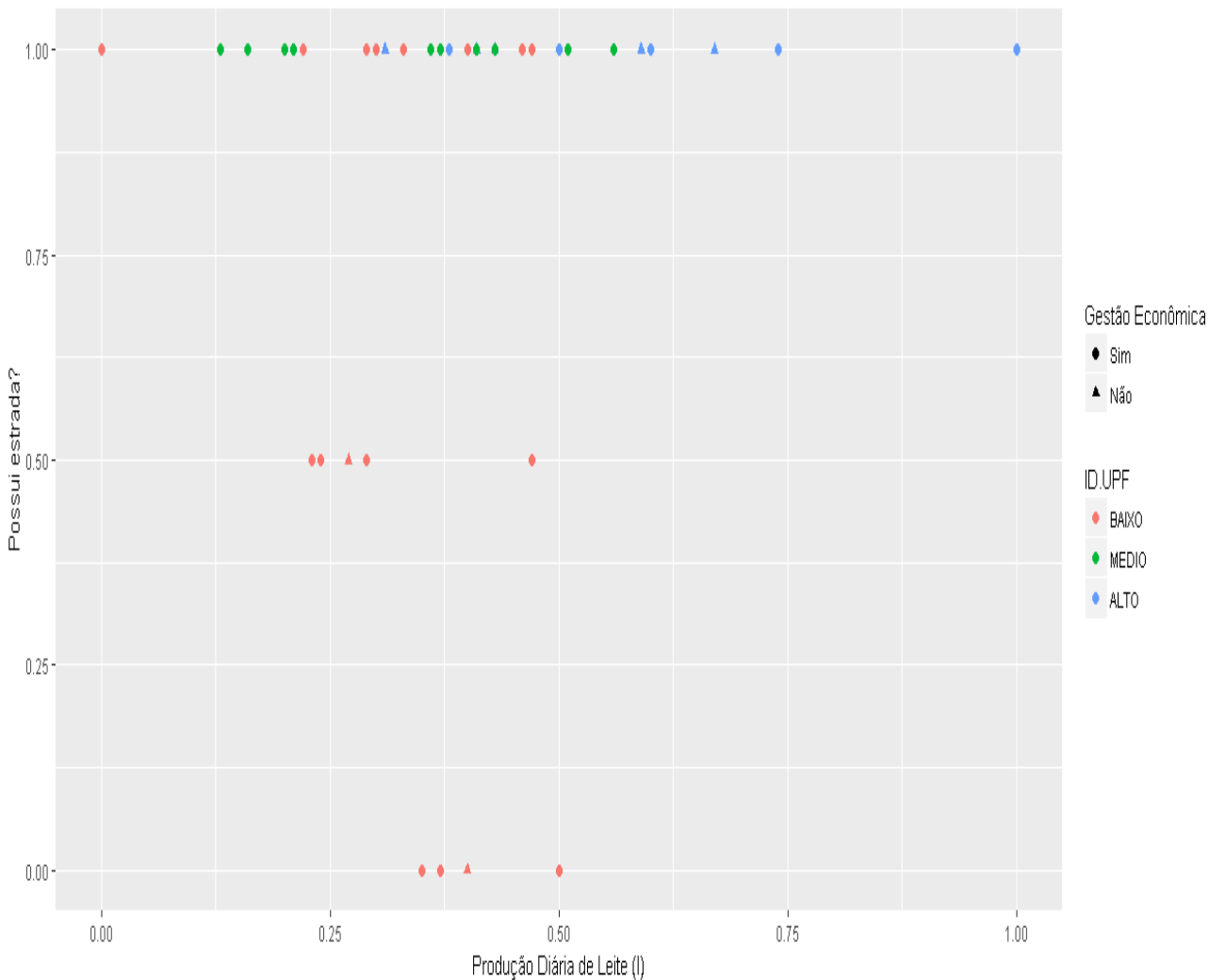
```
n3$F<-cut(n3$F, c(-1,25,50,100,230), labels=c("0-25","26-50","51-100","100-230"))
```

```
ggplot(n3,aes(n3[,4],n3[,5],color=as.factor(n3[,11]),shape=n3[,7]))+  
geom_point(size=2)+  
labs(colour = "Diversidade de Atividades",  
shape="Produção de Leite (l/dia)") +  
scale_x_continuous(name="Receita Anual da Família (R$)") +  
scale_y_continuous(name="Receita Per Capita Mensal (R$)")
```

Passo a Passo



Passo a Passo



```
x2$E2<-cut(x2$E2, c(-1,0.1,1), labels=c("Sim","Não"))
```

```
ggplot(x2,aes(x2$P10,x2$S4,col=x2$ID.UPF,shape=x2$E2))+  
geom_point(size=2)+  
labs(colour = "ID.UPF",shape="Gestão Econômica") +  
scale_x_continuous(name="Produção Diária de Leite (l)") +  
scale_y_continuous(name="Possui estrada?")
```


Referências Bibliográficas

- HAN, J.; KAMBER, M. *Data Mining: Concepts and Techniques*. 2º ed. Morgan Kauf-mann Publishers, p. 5–7, 2006.
- RDocumentation. Disponível em: <<https://www.rdocumentation.org/packages/caret/versions/6.0-76/topics/train>>. Acesso em: 21/05/2017.