

# Atividade (exemplo de relatório) Seleção Estágio TCU

*Augusto Cesar Ribeiro Nunes*

*20 de março de 2015*

## Introdução / Objetivo

Este relatório trata-se ferramenta de análise descritiva de uma amostra da Pesquisa XXXX, com o objetivo de ...

## Metodologia de Amostragem

A amostra foi selecionada de acordo com o critério/técnica ????, a partir de dados da Pesquisa XXXX, publicada no repositório ??? e com Banco de Dados disponível em ????. A base de dados foi devidamente anonimizada.

## Descrição Base de Dados

Composta pelo conjunto de 852 observações das 7 variáveis a seguir:

- **id:** Número do Questionário
- **sexo:** Gênero do respondente, com duas categorias:
  - *Masculino*
  - *Feminino*
- **idade:** Idade do respondente, em anos completos quando da resposta ao questionário (????)
- **localdomicilio:** Local do Domicílio (quanto à classificação da área que ocupa majoritariamente?), com duas categorias:
  - *Urbano*
  - *Rural*
- **especiedomicilio:** Espécie do domicílio, com quatro categorias:
  - *Particular permanente*
  - *Particular improvisado*
  - *Alugado*
  - *Coletivo*
- **qtdcomodos:** Número de cômodos do domicílio
- **pavimentacao:** Se existe calçamento/pavimentação em frente ao domicílio, com três categorias:
  - *Total*
  - *Parcial*
  - *Não existe*

Estrutura da Base de Dados

```
## 'data.frame': 852 obs. of 7 variables:
## $ id : int 1 2 3 4 5 6 7 8 9 10 ...
## $ sexo : int 2 1 2 2 1 2 2 1 1 2 ...
## $ idade : int 75 26 48 73 58 19 21 47 67 71 ...
## $ localdomicilio : int 1 1 2 1 2 1 2 1 1 1 ...
## $ especiedomicilio: int 4 3 4 3 3 3 4 3 3 4 ...
## $ qtdcomodos : int 3 3 1 3 4 1 3 3 4 4 ...
## $ pavimentacao : int 2 1 2 3 1 3 1 1 2 3 ...
```

## Descrição dos respondentes na amostra

### Quanto ao gênero

```
## sexo
##      n missing  unique
##    852      0      2
##
## Masculino (432, 51%), Feminino (420, 49%)
```



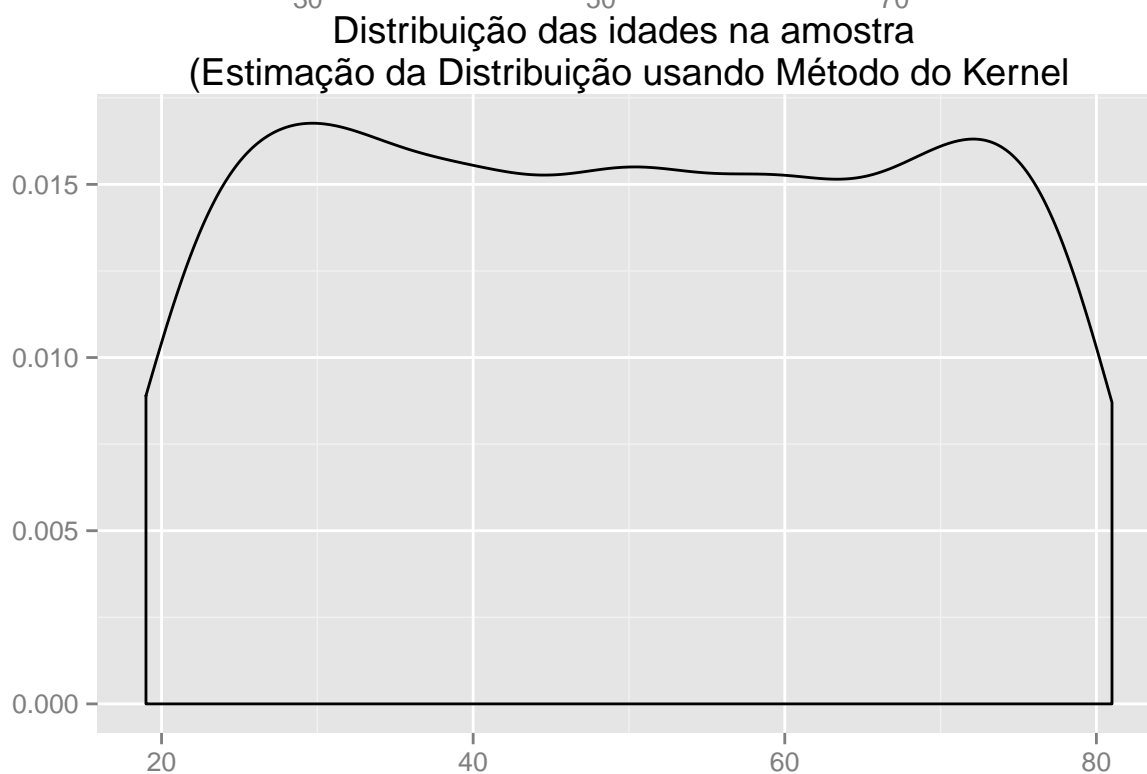
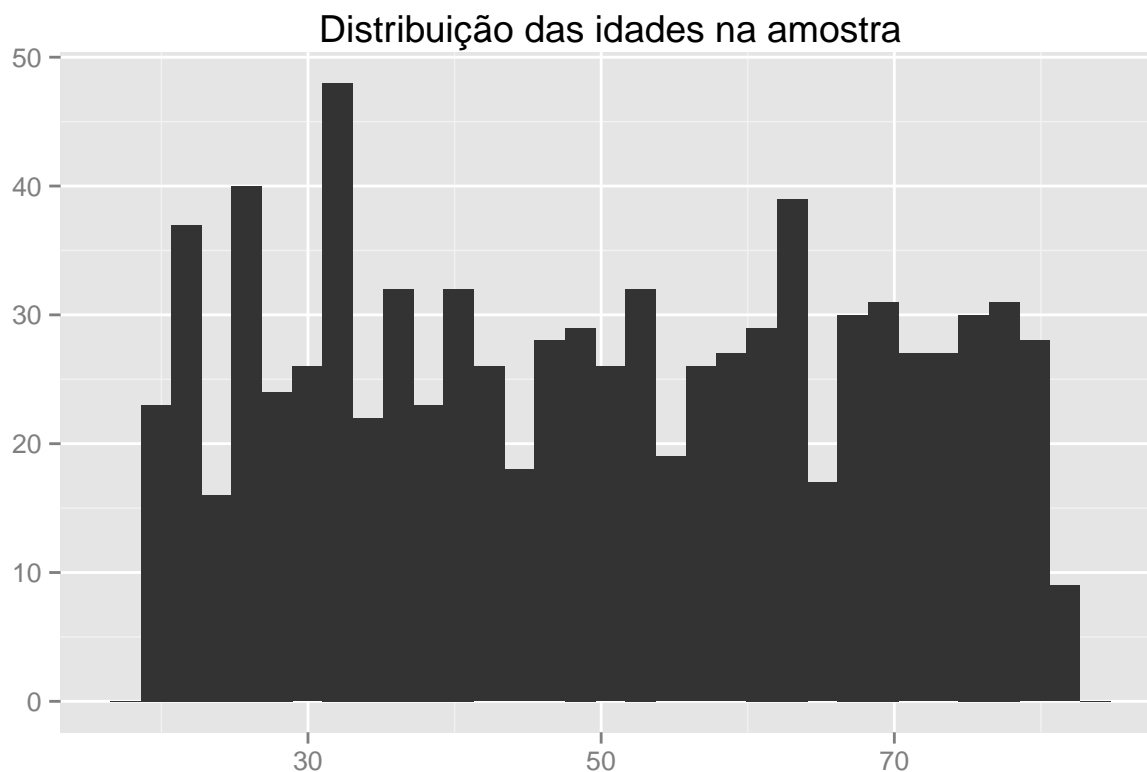
A leve discrepância entre o percentual de homens e mulheres e a relação observada pelo Censo Demográfico, mas nada considerável.

### Quanto à idade

```
## Descrição por idade
##      n missing  unique  Info   Mean   .05   .10   .25   .50
##    852      0      63     1  49.79  21.00  25.00  33.00  50.00
```

```
##      .75      .90      .95
##  66.25  75.00  78.00
##
## lowest : 19 20 21 22 23, highest: 77 78 79 80 81

## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
```

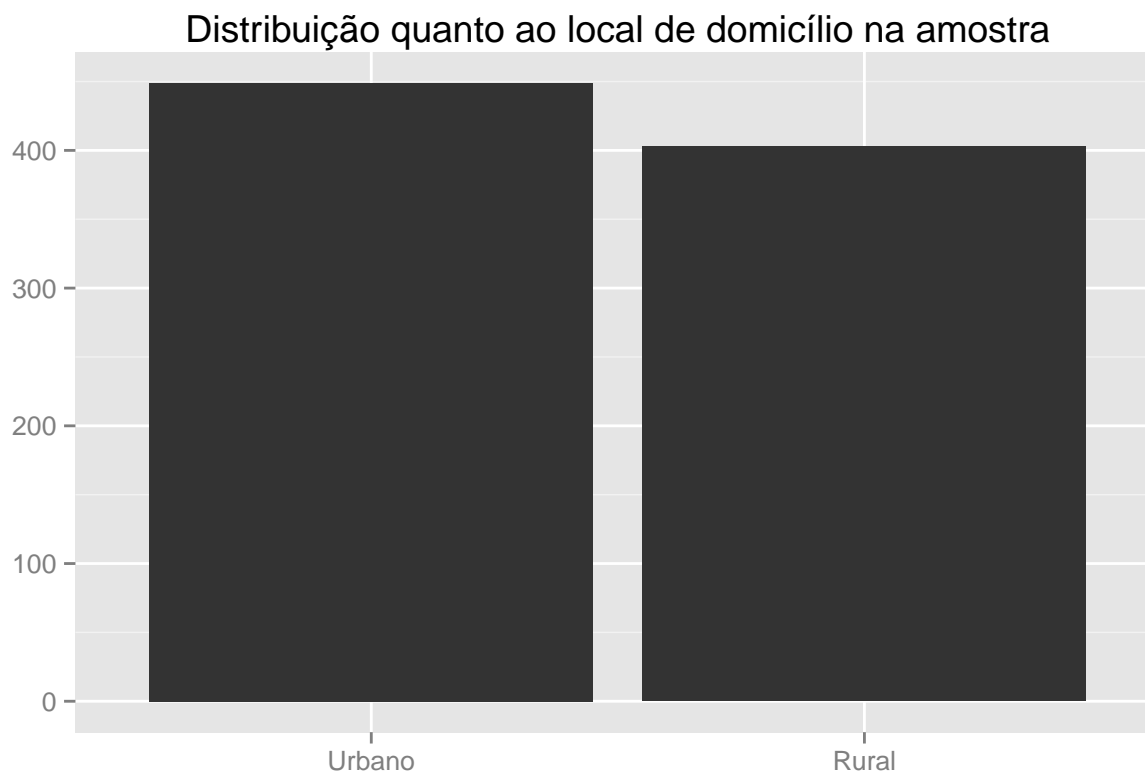


Mediana alta (50 anos), pode ser explicada se essa amostra apresentar um número considerável de proprietários de imóveis e/ou moradores de regiões ou bairros cuja população seja constituída de mais velhos (exemplo de RAs do DF como Lago Sul e Lago Norte).

A amplitude de idades também é grande, 62 anos de idade.

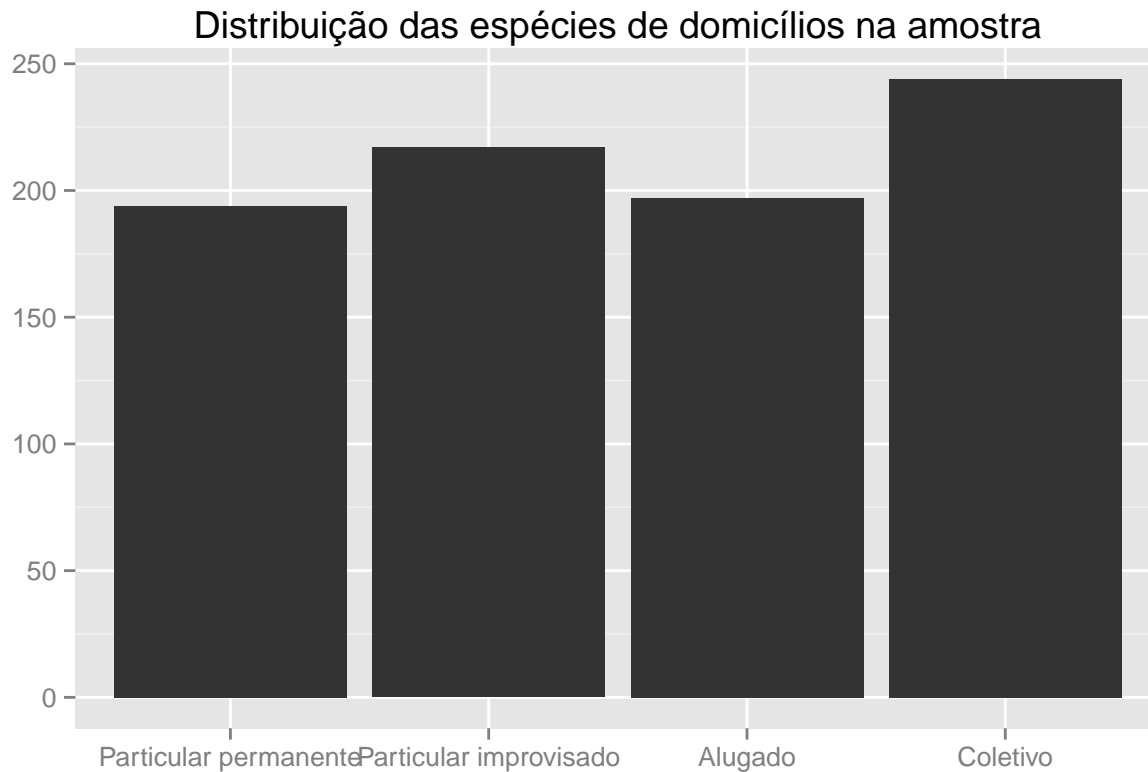
## Quanto ao local de domicílio

```
## Descrição por local do domicílio
##      n missing  unique
##   852      0      2
##
## Urbano (449, 53%), Rural (403, 47%)
```



## Quanto à espécie do domicílio

```
## Descrição por espécie do domicílio
##      n missing  unique
##   852      0      4
##
## Particular permanente (194, 23%)
## Particular improvisado (217, 25%)
## Alugado (197, 23%), Coletivo (244, 29%)
```

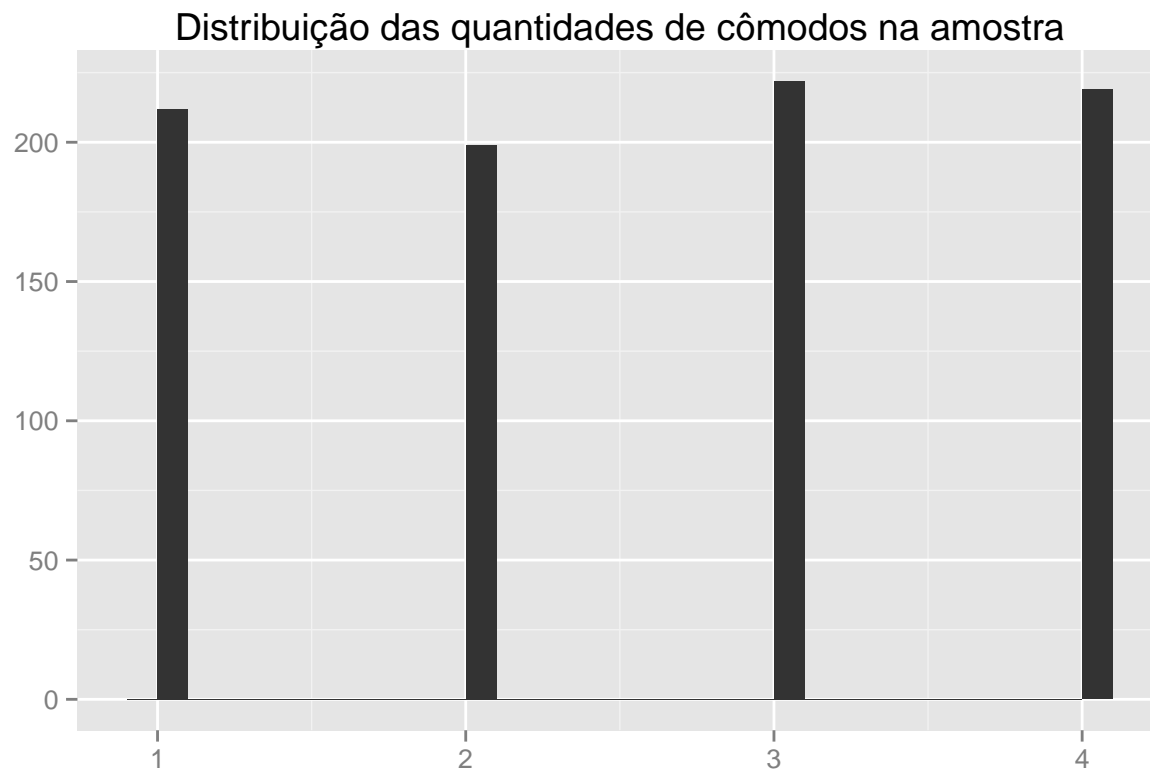


Similarmente ao que ocorreu com as idades, uma discreta diferença entre o percentual de domicílios urbanos e domicílios rurais, o que não é completamente fidedigno à composição por entorno dos domicílios no Brasil: 96,9% dos domicílios particulares permanentes são urbanos. Esta discrepância pode ser explicada por uma diferença de metodologia ou por uma casualidade amostral (Plano Amostral e Planejamento de Pesquisa desconhecidos)

### Quanto à quantidade de cômodos

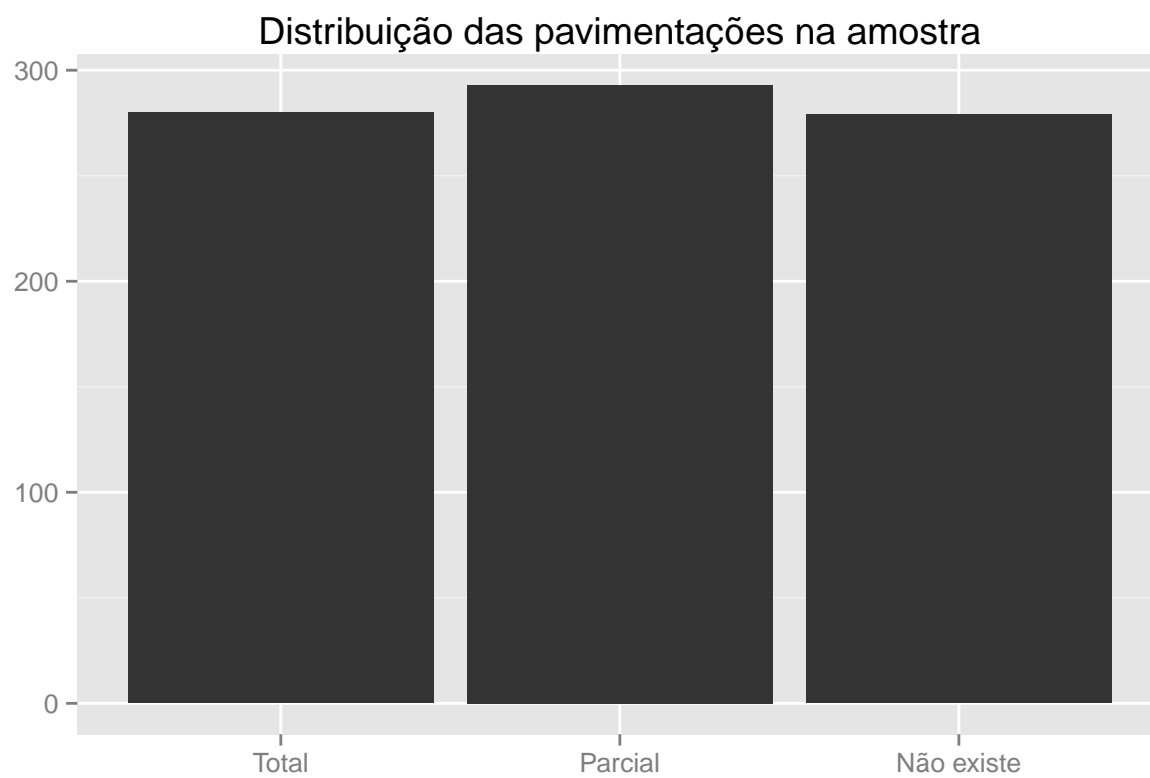
```
## Descrição por quantidade de cômodos
##      n missing  unique    Info   Mean
##    852      0        4    0.94  2.526
##
## 1 (212, 25%), 2 (199, 23%), 3 (222, 26%), 4 (219, 26%)
```

```
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
```



### Quanto à pavimentação

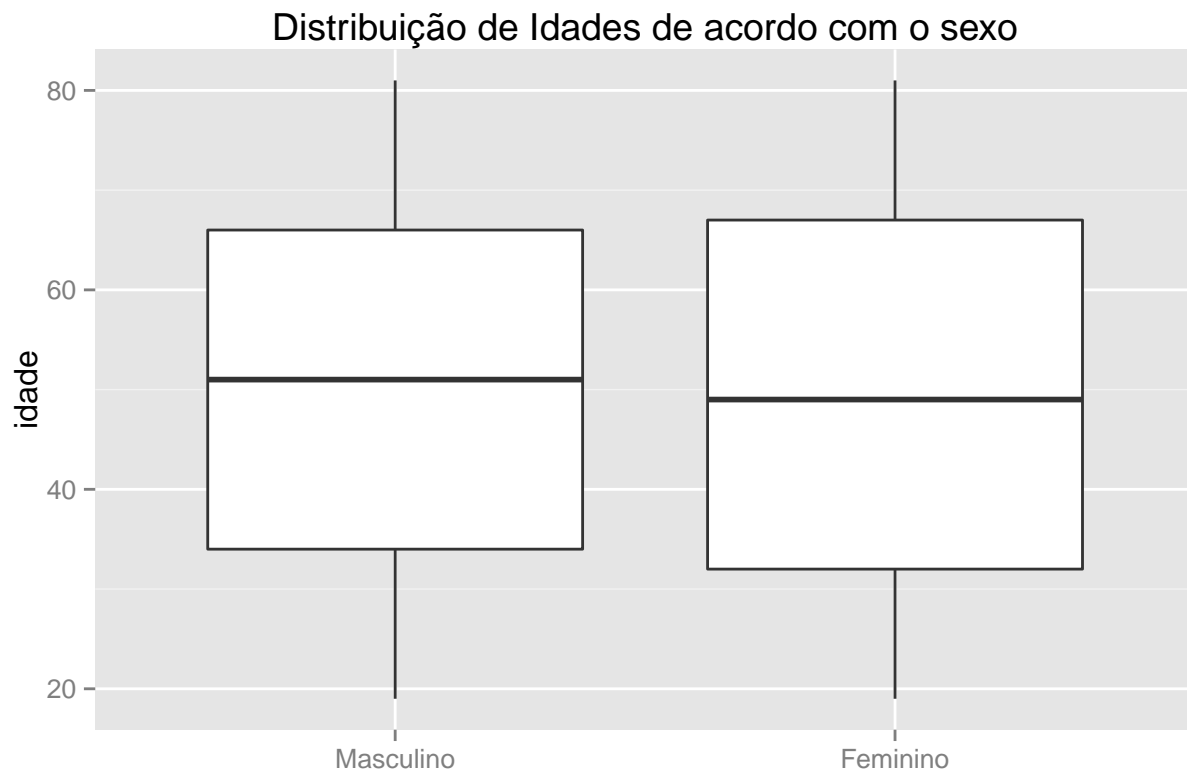
```
## pavimentacao
##      n missing  unique
##    852      0       3
##
## Total (280, 33%), Parcial (293, 34%)
## Não existe (279, 33%)
```





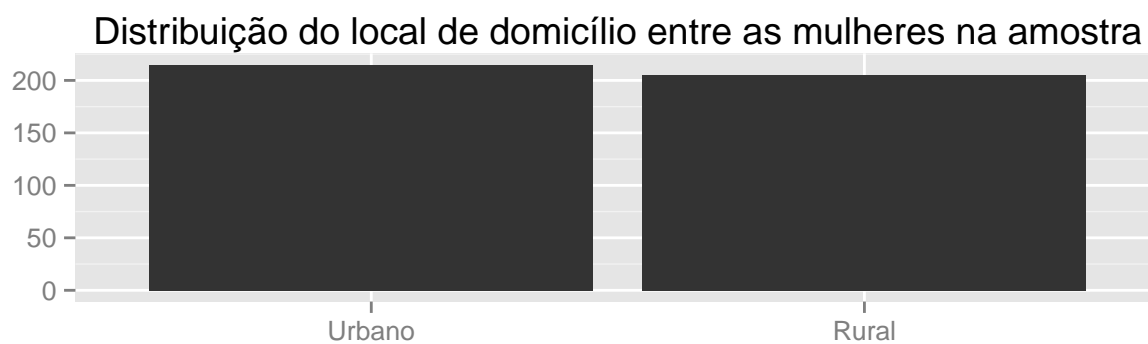
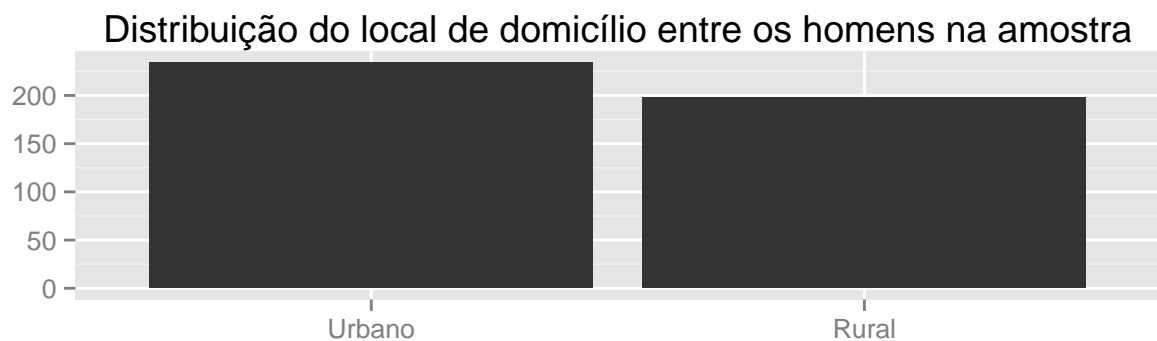
## Análise de relação entre gênero e demais variáveis

### Gênero x Idade



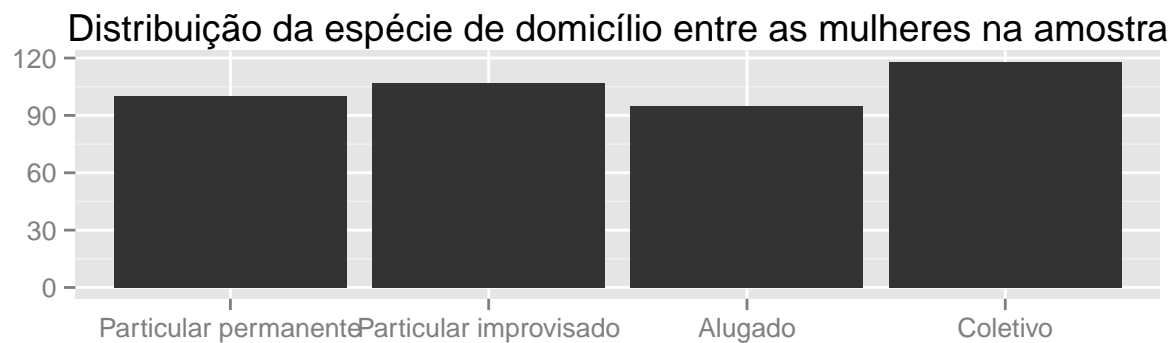
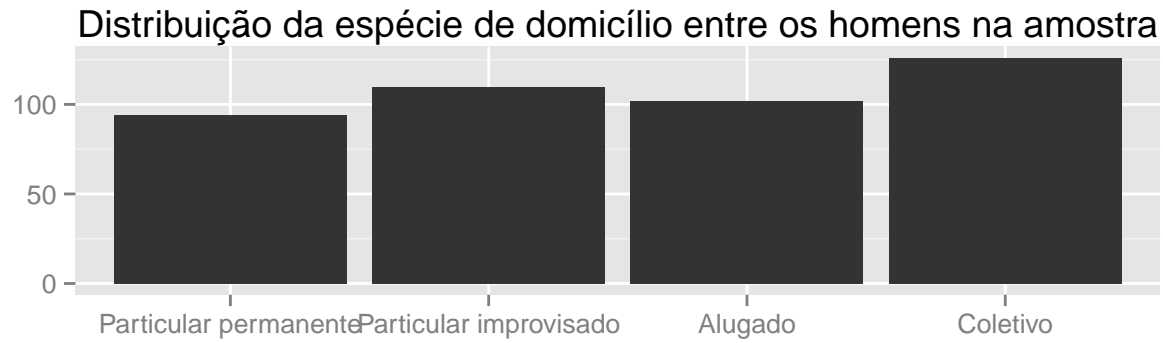
Média de idade dos homens superior à das mulheres, mas não tão marcadamente.

## Gênero x Local do Domicílio



Nenhuma diferença marcante entre o local de domicílios de homens e mulheres respondentes na amostra.

## Gênero x Espécie do Domicílio

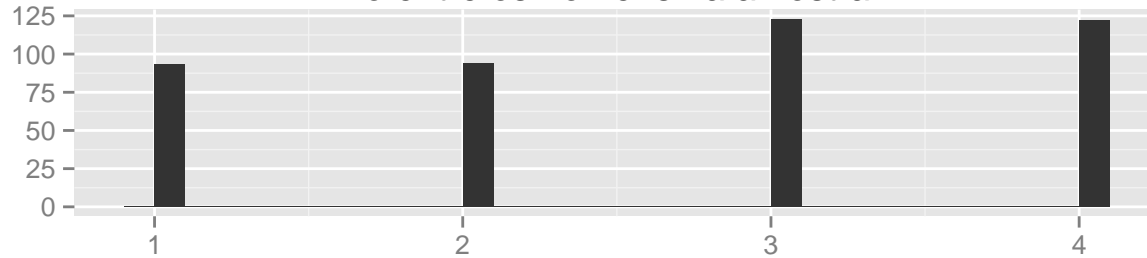


A distribuição de domicílios também não apresenta nada digno de nota quando é discriminada pelo gênero do respondente.

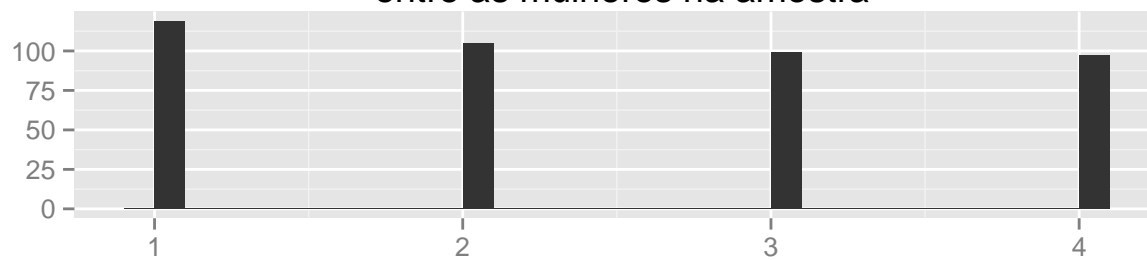
## Gênero x Quantidade de Cômodos

```
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.  
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
```

Distribuição da quantidade de cômodos  
o entre os homens na amostra



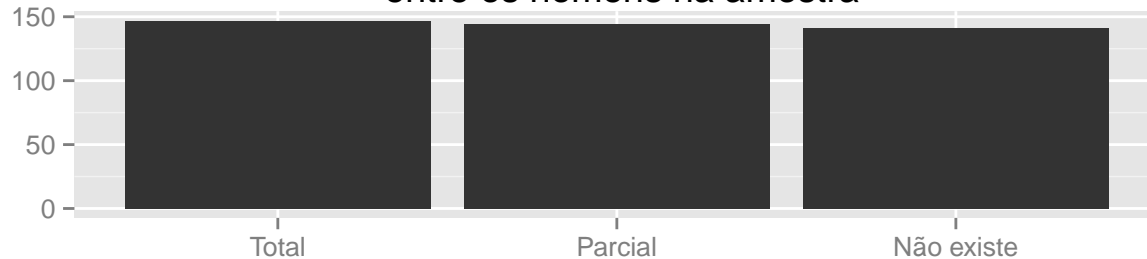
Distribuição da quantidade de cômodos  
entre as mulheres na amostra



Aqui sim uma diferença mais marcante, uma proporção maior de mulheres vive em domicílios com menos cômodos, como se uma distribuição fosse praticamente anti-simétrica à outra.

### Gênero x Pavimentação

Distribuição da qualidade da pavimentação  
entre os homens na amostra



Distribuição da qualidade da pavimentação  
entre as mulheres na amostra



Mais uma vez um déficit qualitativo entre os gêneros. Menos mulheres respondentes vivem em domicílios com pavimentação total.

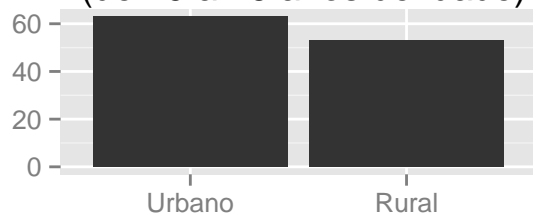
## Análise da relação entre a idade e as demais variáveis

Para enriquecer a análise, usaremos classes de idade da seguinte forma

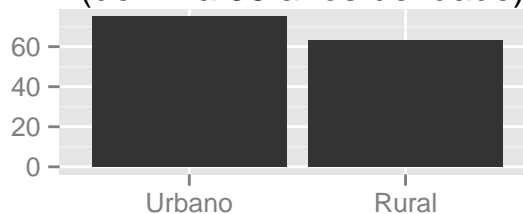
Idade(anos)	Quantidade
19 - 26 anos	19
27 - 36 anos	33
37 - 46 anos	41
47 - 56 anos	48
57 - 66 anos	58
67 - 76 anos	75
76 e mais	81

## Idade x Local do Domicílio

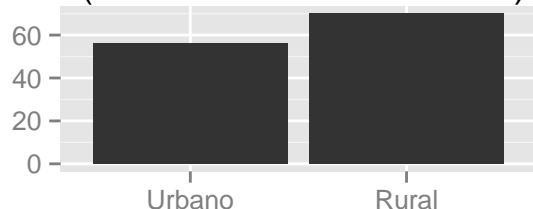
Distribuição local de domicílio  
por classe de idades  
(de 19 a 26 anos de idade)



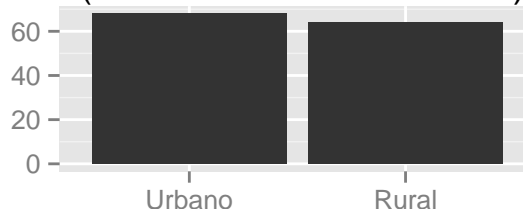
Distribuição local de domicílio  
por classe de idades  
(de 27 a 36 anos de idade)



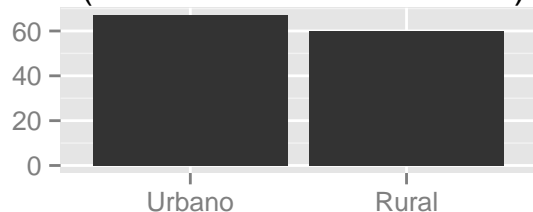
Distribuição local de domicílio  
por classe de idades  
(de 37 a 46 anos de idade)



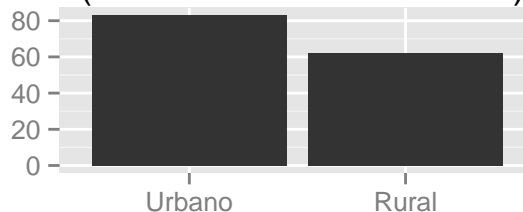
Distribuição do local de domicílio  
por classe de idades  
(de 47 a 56 anos de idade)



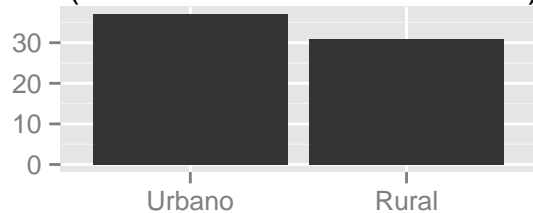
Distribuição do local de domicílio  
por classe de idades  
(de 57 a 66 anos de idade)



Distribuição do local de domicílio  
por classe de idades  
(de 67 a 76 anos de idade)



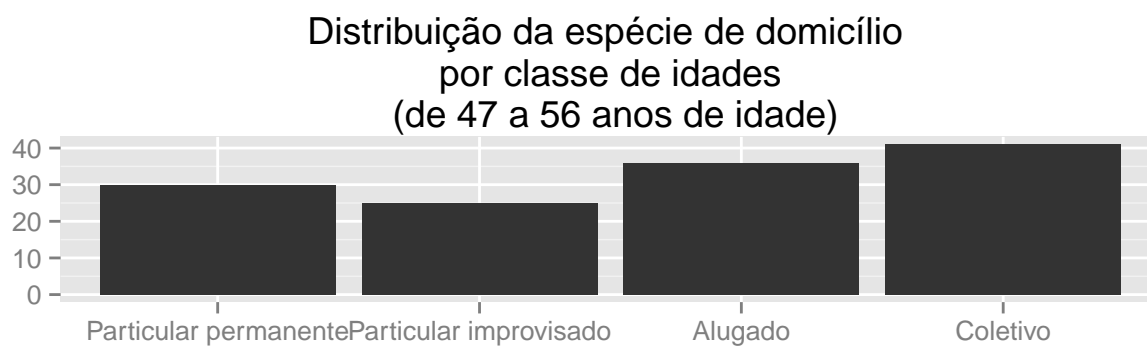
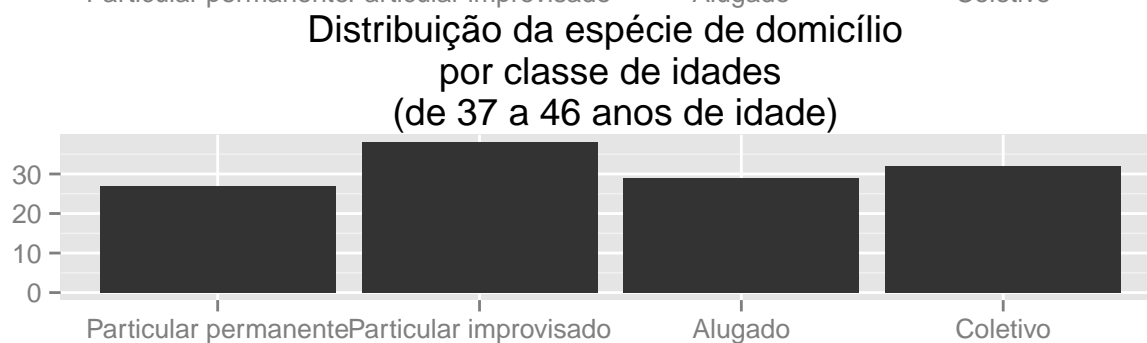
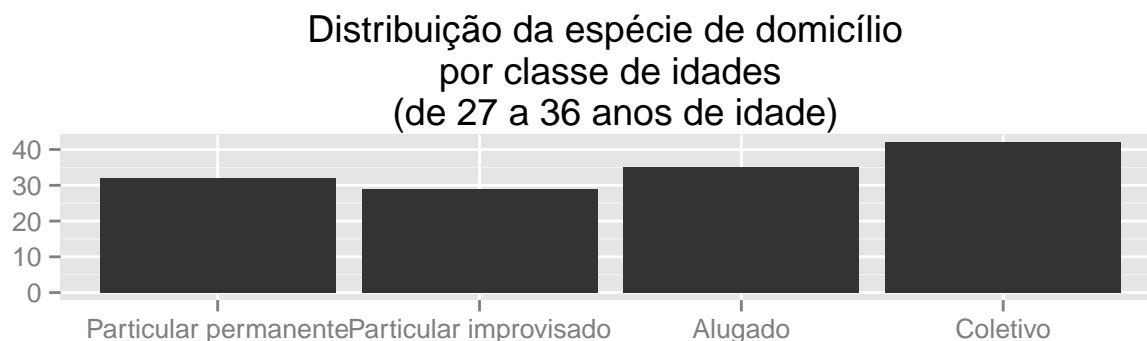
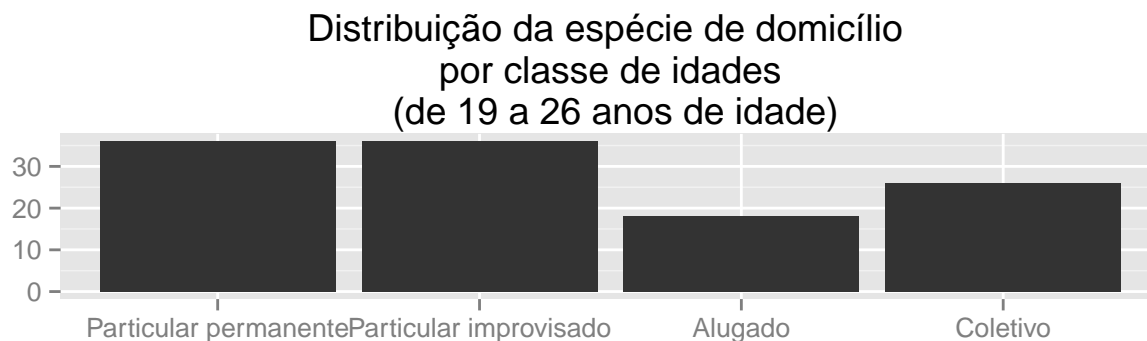
Distribuição local de domicílio  
por classe de idades  
(acima de 77 anos de idade)

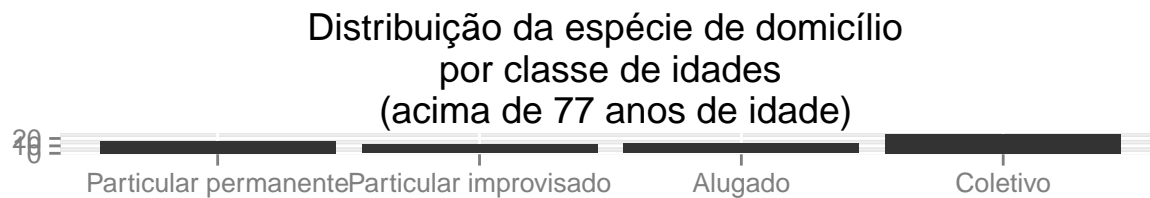
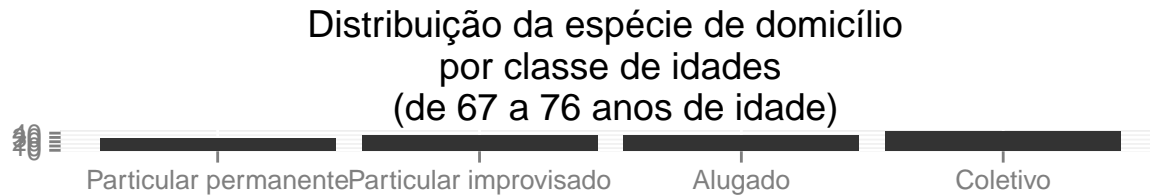
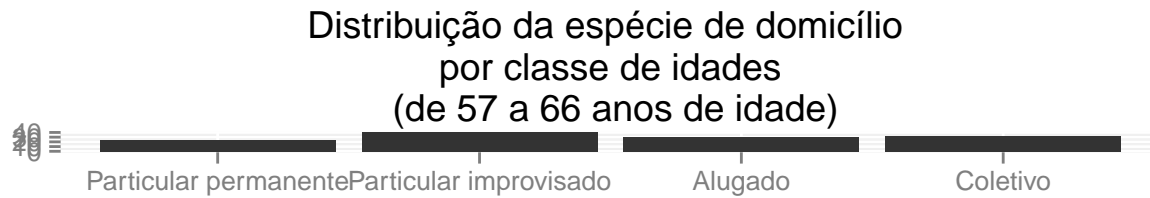


Não há nenhum comportamento sistemático marcante de discriminação do local de moradia por classe de idade, exceto pela classe de 37 a 46 anos de idade, que tem menos pessoas morando em local urbano do que em local rural - ao contrário do que ocorre em todas as outras classes. Pode ser uma mera casualidade amostral, e muito provavelmente é, já que, como foi dito, este cenário não é observado em nenhuma outra

classe de idade.

## Idade x Espécie do Domicílio





Apenas olhando os gráficos já nota-se uma sugestão de tendência nos indivíduos dessa amostra, ao passar dos anos, as pessoas vão migrando de domicílios que se enquadram nas categorias Particular Permanente/Improvisado (de 19 a 26 anos de idade), para Alugado/Coletivo (27 a 36 anos), de volta para domicílios particulares (37 a 46), e no fim da vida, para domicílios Coletivos novamente. É claro que esta é apenas uma especulação dada a estatística descritiva, deveria ser feito um estudo de coortes (longitudinal) para chegar a resultados mais concretos. Mas uma boa indicação. No entanto, o teste  $\chi^2$  nos dá o seguinte resultado:

```
chisq.test(idade,especiedomicilio)
```

```
## Warning in chisq.test(idade, especiedomicilio): Chi-squared approximation
## may be incorrect
```

```
##
## Pearson's Chi-squared test
##
## data: idade and especiedomicilio
## X-squared = 217.68, df = 186, p-value = 0.05573
```

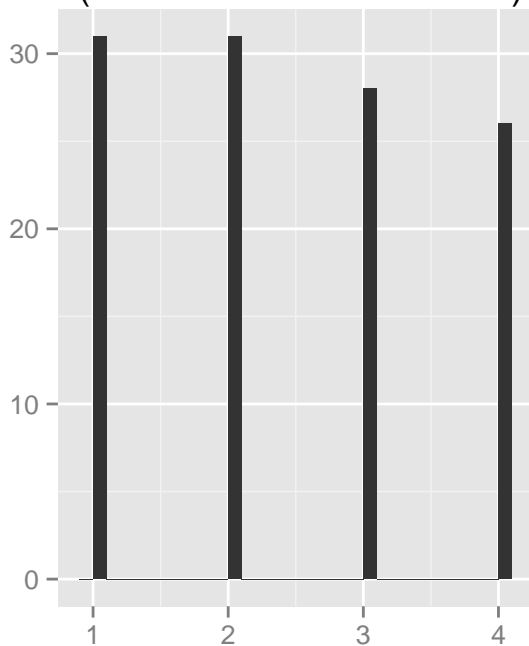
Isto é, é significativa a associação entre a idade e a espécie do domicílio do respondente.

## Idade x Quantidade de Cômodos

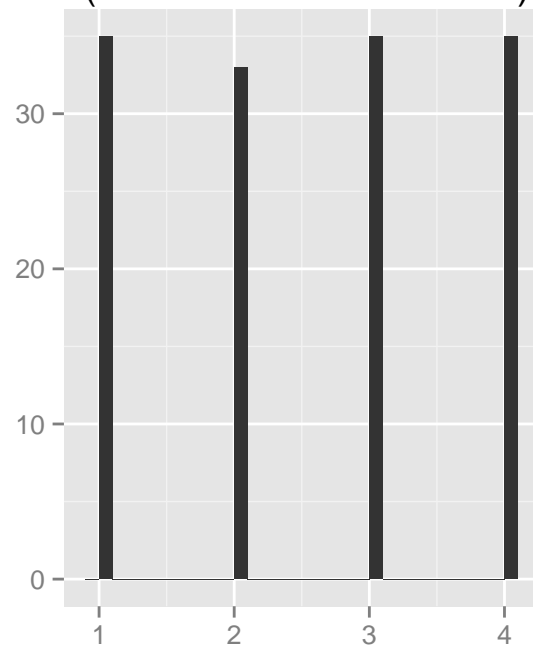
```
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
```



Distribuição da quantidade de cômoc por classe de idades (de 19 a 26 anos de idade)

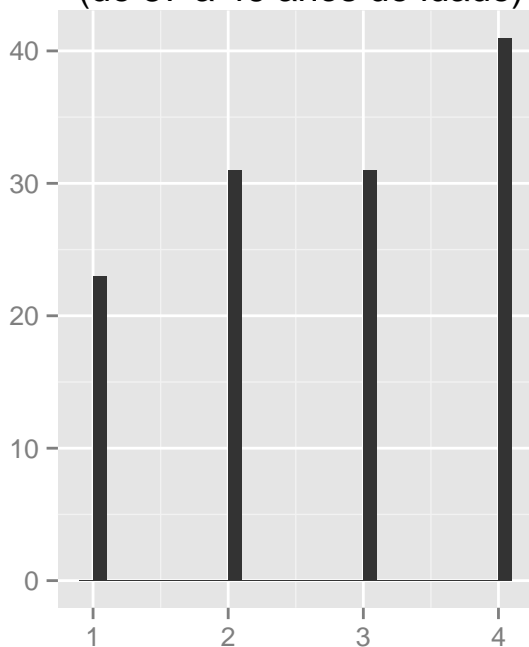


Distribuição da quantidade de cômoc por classe de idades (de 27 a 36 anos de idade)

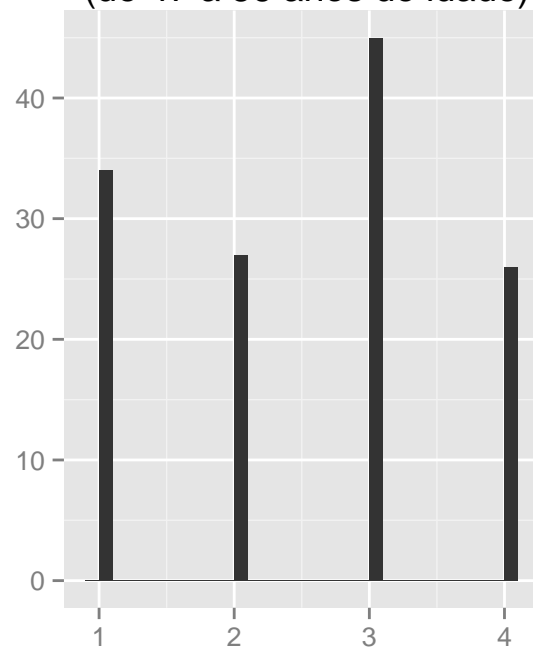


## stat\_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.  
## stat\_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.

Distribuição da quantidade de cômoc por classe de idades (de 37 a 46 anos de idade)



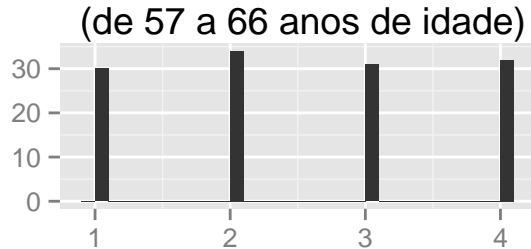
Distribuição da quantidade de cômoc por classe de idades (de 47 a 56 anos de idade)



## stat\_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.

```
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
```

## Distribuição da quantidade de cômodos por classe de idades



## Distribuição da quantidade de cômodos por classe de idades



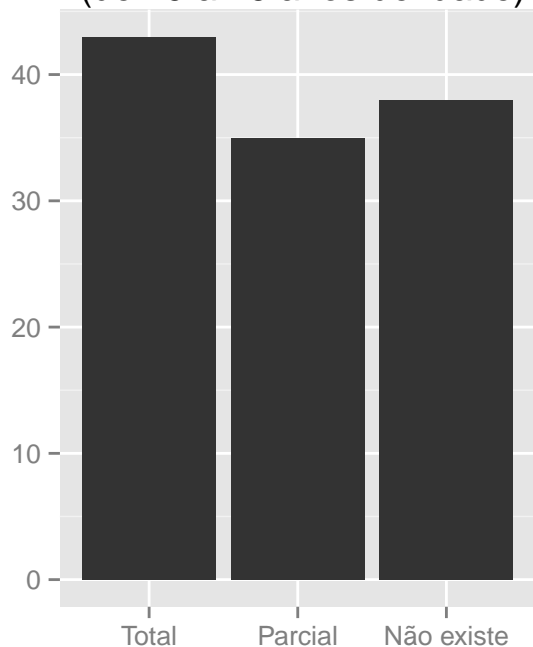
Aqui, a sugestão de tendência é que mais jovens e mais idosos tendem a viver em domicílios com número menor de cômodos. Aqueles que pertencem às classes de idade intermediária, talvez pelo fato de terem estabelecido família - e em particular, terem filhos - precisam de domicílios maiores. Utilizando mais uma vez um teste de hipóteses não-paramétrico - desta vez a *Correlação Produto - Momento de Pearson* - desta vez concluímos que a correlação sugerida pelos gráficos não se sustenta na amostra.

```
cor.test(qtdcomodos,idade)
```

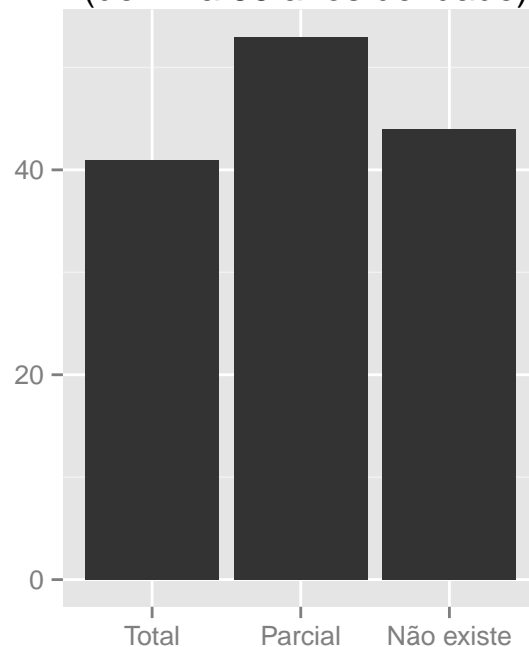
```
##
## Pearson's product-moment correlation
##
## data: qtdcomodos and idade
## t = 0.1669, df = 850, p-value = 0.8675
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## -0.06146291 0.07286174
## sample estimates:
## cor
## 0.005725242
```

## Idade x Pavimentação

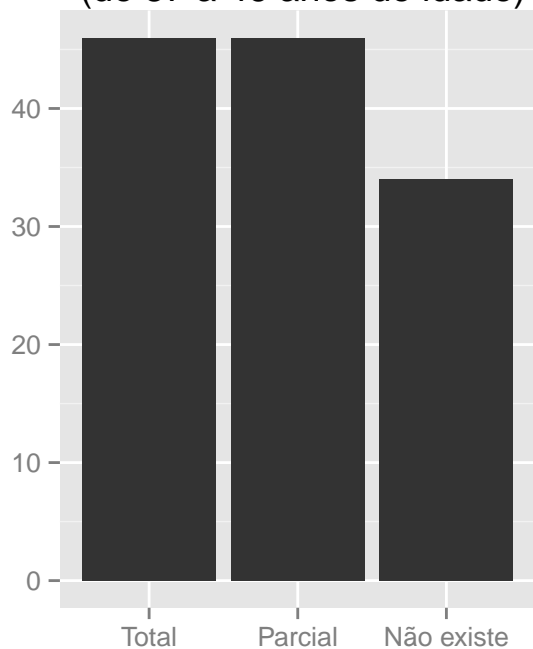
Distribuição da qualidade da pavimentação por classe de idades (de 19 a 26 anos de idade)



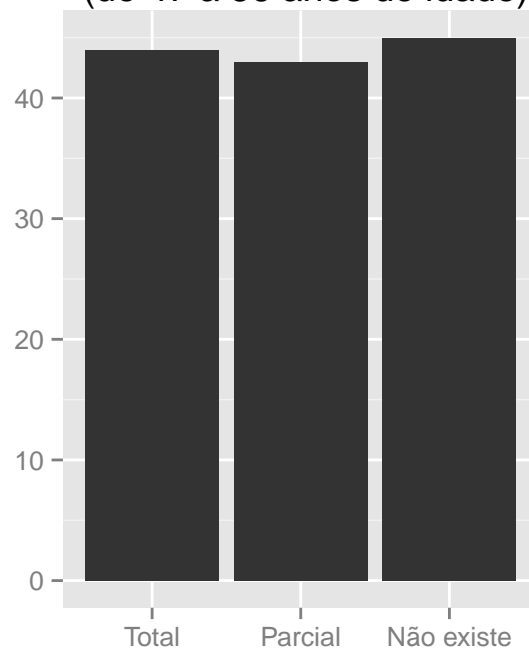
Distribuição da qualidade da pavimentação por classe de idades (de 27 a 36 anos de idade)



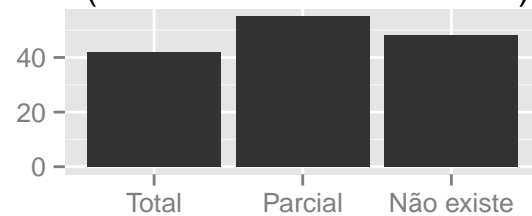
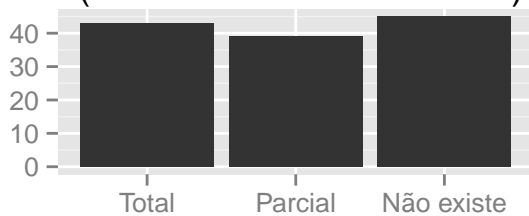
Distribuição da qualidade da pavimentação por classe de idades (de 37 a 46 anos de idade)



Distribuição da qualidade da pavimentação por classe de idades (de 47 a 56 anos de idade)



Distribuição da qualidade da pavimentação por classe de idades (de 57 a 66 anos de idade)



Distribuição da qualidade da pavimentação por classe de idades (acima de 77 anos de idade)



Utilizando novamente o teste  $\chi^2$ , concluímos que não há significância considerável na relação entre a idade e o tipo de pavimentação.

```
chisq.test(idade, pavimentacao)
```

```
## Warning in chisq.test(idade, pavimentacao): Chi-squared approximation may
## be incorrect
```

```
##
## Pearson's Chi-squared test
##
## data: idade and pavimentacao
## X-squared = 125.0989, df = 124, p-value = 0.4555
```

## Análises de Regressão

A Idade influencia na espécie e local do domicílio?

```
##
## Call:
## lm(formula = idade ~ localdomicilio + especiedomicilio)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -32.700 -16.151  -0.475  16.137  33.168
##
## Coefficients:
```

```
##                                Estimate Std. Error t value
## (Intercept)                   48.5553      1.4759  32.899
## localdomicilioRural          -0.7228      1.2630  -0.572
## especiedomicilioParticular improvisado  0.5628      1.8146   0.310
## especiedomicilioAlugado        2.3181      1.8589   1.247
## especiedomicilioColetivo        3.1451      1.7701   1.777
##                                Pr(>|t|)
## (Intercept)                   <2e-16 ***
## localdomicilioRural           0.567
## especiedomicilioParticular improvisado  0.757
## especiedomicilioAlugado        0.213
## especiedomicilioColetivo        0.076 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 18.36 on 847 degrees of freedom
## Multiple R-squared:  0.005505, Adjusted R-squared:  0.0008081
## F-statistic: 1.172 on 4 and 847 DF, p-value: 0.3216
```

A única relação com discreta significância é entre a Idade e o Domicílio Coletivo.

## A Idade influencia no tipo de pavimentação e no número de cômodos?

```
##
## Call:
## lm(formula = idade ~ pavimentacao + qtdcomodos)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -31.492 -16.516  -0.026   16.436   32.205
##
## Coefficients:
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   48.60262      1.79418  27.089  <2e-16 ***
## pavimentacaoParcial           1.23071      1.53654   0.801   0.423
## pavimentacaoNão existe        1.60025      1.55529   1.029   0.304
## qtdcomodos                     0.09642      0.56095   0.172   0.864
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 18.39 on 848 degrees of freedom
## Multiple R-squared:  0.001404, Adjusted R-squared: -0.002128
## F-statistic: 0.3976 on 3 and 848 DF, p-value: 0.7548
```

O resultado da regressão nos diz que não há significância no relacionamento entre estas três variáveis.

## O gênero influencia no tipo de pavimentação, local da moradia, ou número de cômodos?

```
##
## Call:
## lm(formula = dados$sexo ~ pavimentacao + localdomicilio + qtdcomodos)
```

```
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.5900 -0.4843 -0.3971  0.4971  0.6029
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.57139    0.05143  30.552 < 2e-16 ***
## pavimentacaoParcial  0.03338    0.04168   0.801  0.42342
## pavimentacaoNão existe 0.01991    0.04221   0.472  0.63733
## localdomicilioRural  0.02880    0.03427   0.840  0.40106
## qtdcomodos       -0.04357    0.01521  -2.864  0.00429 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4986 on 847 degrees of freedom
## Multiple R-squared:  0.0112, Adjusted R-squared:  0.00653
## F-statistic: 2.398 on 4 and 847 DF,  p-value: 0.04871
```

O sexo influencia significativamente no número de cômodos, mulheres residem em moradias com menos cômodos.