

Caso: Análise da distribuição de equipamentos culturais na cidade de São Paulo

```
library(dplyr)

##
## Attaching package: 'dplyr'
##
## The following objects are masked from 'package:stats':
##
##   filter, lag
##
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

library(sf)

## Linking to GEOS 3.8.0, GDAL 3.0.4, PROJ 6.3.1

library(tidyr)
library(purrr)
library(ggplot2)
```

Importar os dados

Cultura

```
equipamentos_cultura <- c("bibliotecas", "espacos_culturais", "museus", "outros", "teatro_cinema")
arquivos <- list.files("../dados/cultura-raw")
shapes_cultura <- arquivos[grep(".*SIRGAS.*CULTURA.*shp", arquivos)]

cultura <- map2(
  shapes_cultura, equipamentos_cultura,
  ~read_sf(paste0("../dados/cultura-raw/", .x), options = "ENCODING=windows-1252") %>%
    mutate(tipo = .y) %>%
    select(everything(), geometry)
) %>%
do.call(rbind, .)
```

Distritos

```
distritos <- read_sf("../dados/distritos-raw/35DSE250GC_SIR.shp", options = "ENCODING=windows-1252")
```

Preparação dos dados

Distritos

```

distritos <- distritos %>%
  st_transform(31983) %>%
  rename(cod_distrito = CD_GEOCODD, nome_distrito = NM_DISTRIT) %>%
  filter(grepl(3550308, cod_distrito)) %>%
  mutate(area_distrito = as.numeric(st_area(geometry)) / 106) %>%
  select(cod_distrito, nome_distrito, area_distrito)

```

Mesclar dados

```

cultura <- cultura %>%
  st_set_crs(31983) %>%
  st_join(distritos, join = st_intersects) %>%
  as_tibble() %>%
  count(cod_distrito, tipo) %>%
  spread(tipo, n) %>%
  mutate_all(~replace_na(., 0)) %>%
  mutate(cod_distrito = as.character(cod_distrito)) %>%
  rbind(data.frame(cod_distrito = distritos$cod_distrito[!(distritos$cod_distrito %in% .$cod_distrito)]
    , bibliotecas = rep(0, 5),
    espacos_culturais = rep(0, 5),
    museus = rep(0, 5),
    outros = rep(0, 5),
    teatro_cinema = rep(0, 5)))

cultura <- cultura %>%
  mutate(equipamentos = bibliotecas + teatro_cinema + outros + espacos_culturais + museus) %>%
  merge(select(distritos, cod_distrito, nome_distrito), by = "cod_distrito")

cultura <- st_as_sf(cultura)

```

Dados preparados

```
head(data.frame(cultura))
```

```

##   cod_distrito bibliotecas espacos_culturais museus outros teatro_cinema
## 1   355030801           2              0      0      0              0
## 2   355030802           2              2      1      0             11
## 3   355030803           3              0      0      0              1
## 4   355030804           1              0      0      0              1
## 5   355030805           1              0      0      0              1
## 6   355030806           0              3      2      0             31
##   equipamentos      nome_distrito      geometry
## 1             2      ÁGUA RASA MULTIPOLYGON (((338561.8 73...
## 2            16 ALTO DE PINHEIROS MULTIPOLYGON (((325006.5 73...
## 3             4      ANHANGUERA MULTIPOLYGON (((313586.5 74...
## 4             2      ARICANDUVA MULTIPOLYGON (((344972.5 73...
## 5             2      ARTUR ALVIM MULTIPOLYGON (((349341.7 73...
## 6            36      BARRA FUNDA MULTIPOLYGON (((327682.2 73...

```

Análise exploratória

Sumário estatístico

```
equipamentos <- cultura %>%  
  data.frame %>%  
  select_if(is.numeric) %>%  
  pivot_longer(cols=everything())
```

equipamentos

```
## # A tibble: 576 x 2  
##   name      value  
##   <chr>    <dbl>  
## 1 bibliotecas      2  
## 2 espacos_culturais 0  
## 3 museus           0  
## 4 outros           0  
## 5 teatro_cinema     0  
## 6 equipamentos      2  
## 7 bibliotecas      2  
## 8 espacos_culturais 2  
## 9 museus           1  
## 10 outros          0  
## # ... with 566 more rows
```

```
sumario <- equipamentos %>%  
  group_by(name) %>%  
  summarise_at(vars(value), funs(sum, min, mean, median, max, IQR, sd))
```

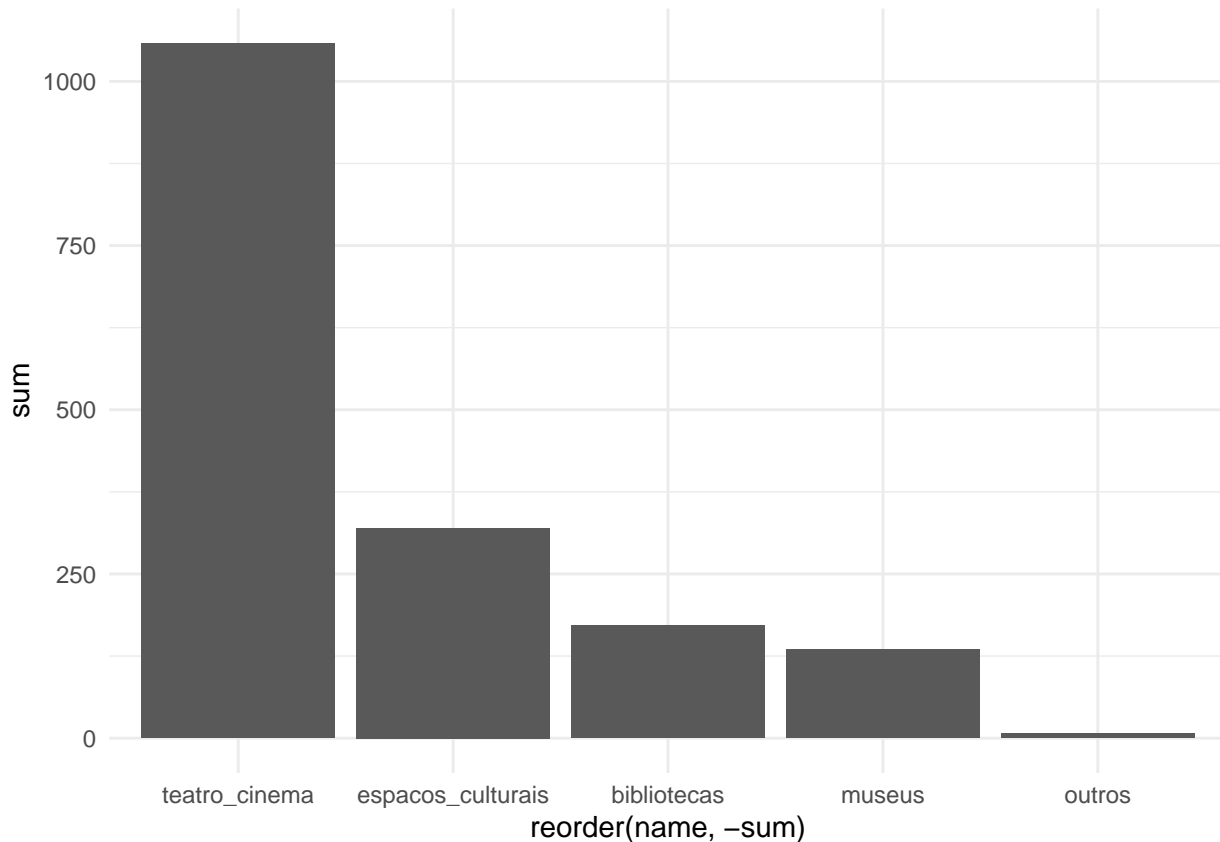
```
## Warning: 'funs()' is deprecated as of dplyr 0.8.0.  
## Please use a list of either functions or lambdas:  
##  
## # Simple named list:  
##   list(mean = mean, median = median)  
##  
## # Auto named with 'tibble::lst()':  
##   tibble::lst(mean, median)  
##  
## # Using lambdas  
##   list(~ mean(., trim = .2), ~ median(., na.rm = TRUE))  
## This warning is displayed once every 8 hours.  
## Call 'lifecycle::last_warnings()' to see where this warning was generated.
```

sumario

```
## # A tibble: 6 x 8  
##   name      sum  min  mean median  max  IQR  sd  
##   <chr>    <dbl> <dbl>  <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 bibliotecas      171    0  1.78      1    13    1   1.96  
## 2 equipamentos    1691    0 17.6      8   174 12.2 29.2  
## 3 espacos_culturais  320    0  3.33      1    64    2   9.55  
## 4 museus          135    0  1.41      0    20    1   3.04  
## 5 outros           7    0 0.0729     0     3    0   0.363  
## 6 teatro_cinema   1058    0 11.0      3   104   10  19.3
```

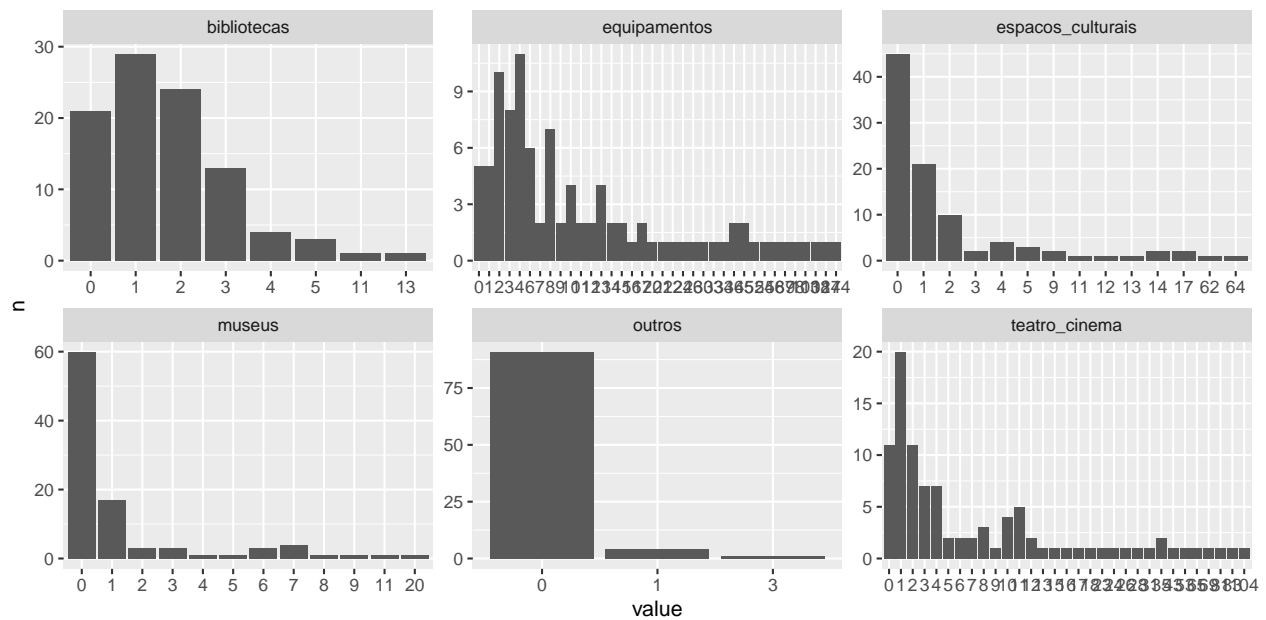
Quantidades por tipo de equipamento

```
library(ggplot2)
sumario %>%
  filter(name != "equipamentos") %>%
  ggplot(aes(x=reorder(name, -sum), y=sum)) +
  geom_bar(stat='identity') +
  theme_minimal()
```



Qual a distribuição de cada equipamento pela cidade?

```
equipamentos %>%
  group_by(name) %>%
  mutate_at(vars(-name), ~factor(., levels = seq(0, max(.), 1))) %>%
  count(name, value) %>%
  ggplot(aes(x=value, y=n)) +
  geom_bar(stat="identity") +
  facet_wrap(~name, scales = "free")
```



Distribuição territorial dos equipamentos

```
equips_distritos <- cultura %>%
  select(-cod_distrito)
```

```
equips_distritos
```

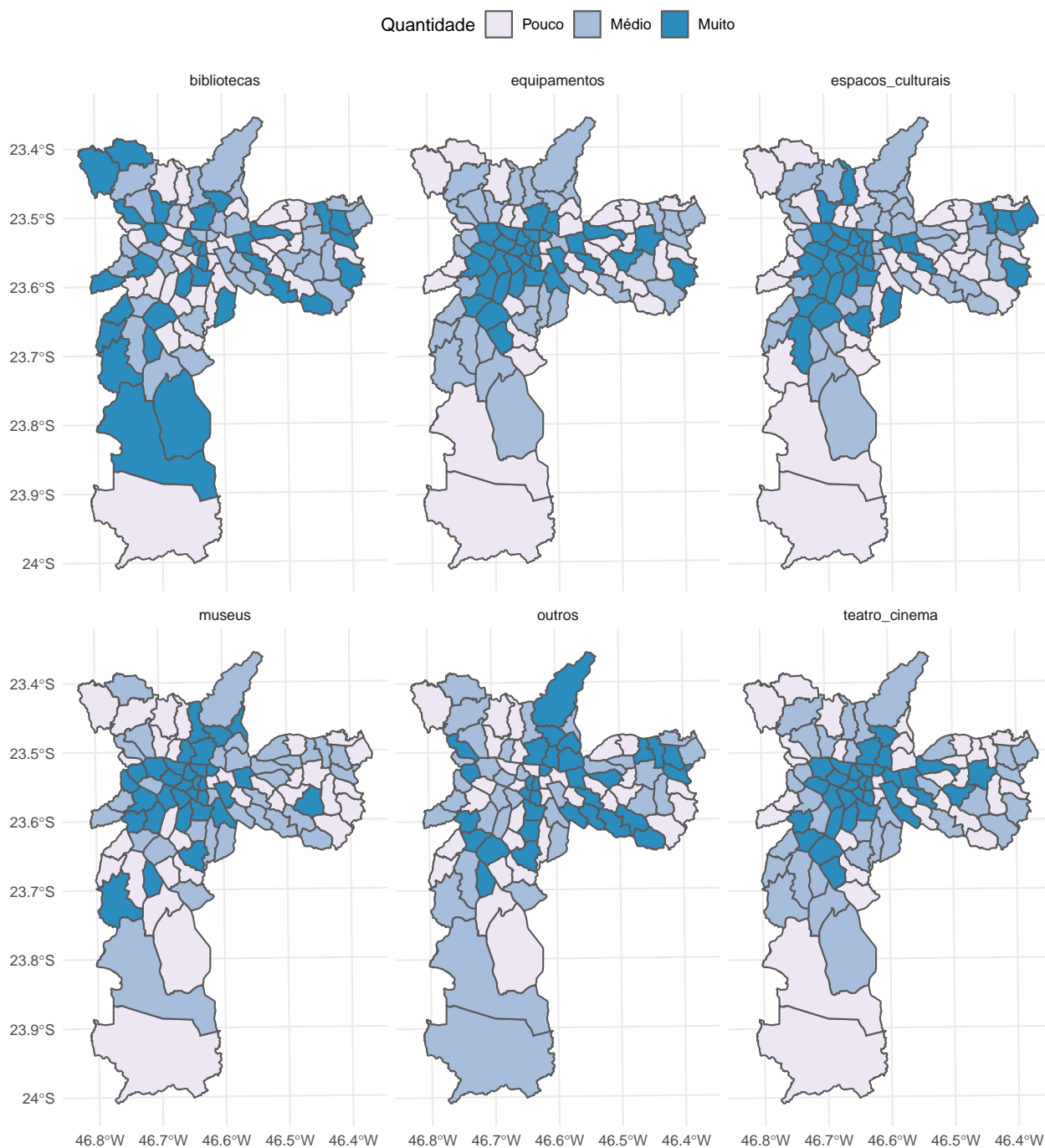
```
## Simple feature collection with 96 features and 7 fields
## geometry type:  MULTIPOLYGON
## dimension:      XY
## bbox:           xmin: 313394.9 ymin: 7343708 xmax: 360619.8 ymax: 7416189
## projected CRS:  SIRGAS 2000 / UTM zone 23S
## First 10 features:
##      bibliotecas espacos_culturais museus outros teatro_cinema equipamentos
## 1              2                  0      0      0              0              2
## 2              2                  2      1      0             11             16
## 3              3                  0      0      0              1              4
## 4              1                  0      0      0              1              2
## 5              1                  0      0      0              1              2
## 6              0                  3      2      0             31             36
## 7              0                 13      6      1             83            103
## 8              0                  2      0      0              9             11
## 9              2                  2      7      0             10             21
## 10             0                  1      0      0              0              1
##      nome_distrito geometry
## 1      ÁGUA RASA MULTIPOLYGON (((338561.8 73...
## 2    ALTO DE PINHEIROS MULTIPOLYGON (((325006.5 73...
## 3      ANHANGUERA MULTIPOLYGON (((313586.5 74...
## 4      ARICANDUVA MULTIPOLYGON (((344972.5 73...
## 5    ARTUR ALVIM MULTIPOLYGON (((349341.7 73...
## 6    BARRA FUNDA MULTIPOLYGON (((327682.2 73...
## 7      BELA VISTA MULTIPOLYGON (((331873.4 73...
## 8          BELÉM MULTIPOLYGON (((336579.3 73...
## 9      BOM RETIRO MULTIPOLYGON (((334062.6 73...
```

```
## 10          BRÁS MULTIPOLYGON (((335159.1 73...
```

```
equips_distritos %>%  
  mutate_if(is.numeric, ~ntile(., 3)) %>%  
  mutate_if(is.numeric, ~factor(., labels = c("Pouco", "Médio", "Muito"))) %>%  
  pivot_longer(cols=c(1:6)) %>%  
  st_as_sf() %>%  
  ggplot(aes(fill=value)) +  
  geom_sf() +  
  scale_fill_brewer(palette="PuBu", "Quantidade") +  
  theme_minimal() +  
  facet_wrap(~name) +  
  theme(legend.position = "top") +  
  labs(title = "Distribuição dos equipamentos culturais na cidade de São Paulo")
```

```
## Warning in val_cols[col_id] <- unname(as.list(data[cols])): number of items to  
## replace is not a multiple of replacement length
```

Distribuição dos equipamentos culturais na cidade de São Paulo



Distribuição da população

```
censo <- read.csv("../dados/censo.csv") %>%
  mutate(habitantes = indigenas + pretos + pardos + brancos + amarelos,
         renda_media_dom = renda_domicilios / domicilios,
         renda_pc = renda_domicilios / habitantes)

cultura <- cultura %>%
```

```

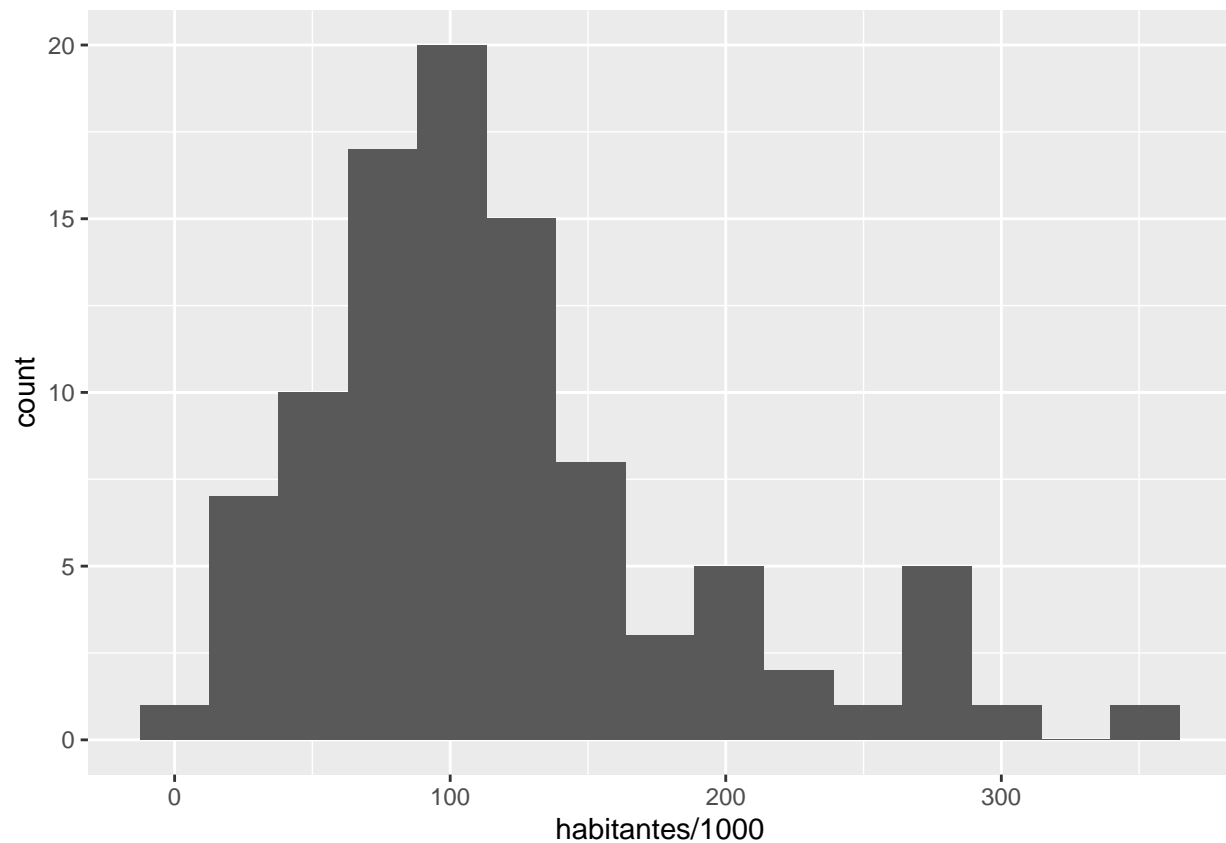
merge(censo, by = "cod_distrito")

glimpse(cultura)

## Rows: 96
## Columns: 19
## $ cod_distrito      <chr> "355030801", "355030802", "355030803", "355030804...
## $ bibliotecas       <dbl> 2, 2, 3, 1, 1, 0, 0, 0, 2, 0, 1, 4, 1, 0, 0, 0, 4...
## $ espacos_culturais <dbl> 0, 2, 0, 0, 0, 3, 13, 2, 2, 1, 1, 4, 3, 1, 1, 1, ...
## $ museus            <dbl> 0, 1, 0, 0, 0, 2, 6, 0, 7, 0, 0, 20, 0, 1, 0, 0, ...
## $ outros            <dbl> 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0...
## $ teatro_cinema     <dbl> 0, 11, 1, 1, 1, 31, 83, 9, 10, 0, 2, 6, 4, 1, 1, ...
## $ equipamentos     <dbl> 2, 16, 4, 2, 2, 36, 103, 11, 21, 1, 4, 34, 8, 3, ...
## $ nome_distrito     <chr> "ÁGUA RASA", "ALTO DE PINHEIROS", "ANHANGUERA", "...
## $ brancos           <int> 71095, 38262, 32321, 62365, 64722, 11640, 51075, ...
## $ pretos            <int> 2154, 860, 4282, 4914, 10406, 438, 3377, 1777, 15...
## $ amarelos          <int> 2087, 1361, 333, 2532, 1452, 417, 3253, 1192, 514...
## $ pardos            <int> 9565, 2615, 28783, 19762, 28622, 1811, 11659, 935...
## $ indigenas         <int> 62, 19, 70, 49, 67, 17, 96, 189, 138, 123, 261, 5...
## $ domicilios        <int> 28652, 15408, 19437, 27661, 33712, 5623, 29967, 1...
## $ renda_domicilios <int> 114911602, 170039733, 34768395, 71198307, 7788354...
## $ habitantes        <int> 84963, 43117, 65789, 89622, 105269, 14323, 69460,...
## $ renda_media_dom   <dbl> 4010.596, 11035.808, 1788.774, 2573.960, 2310.262...
## $ renda_pc          <dbl> 1352.4899, 3943.6819, 528.4834, 794.4289, 739.852...
## $ geometry          <MULTIPOLYGON [m]> MULTIPOLYGON (((338561.8 73..., MULT...

ggplot(cultura, aes(x=habitantes / 1000)) +
  geom_histogram(bins = 15)

```

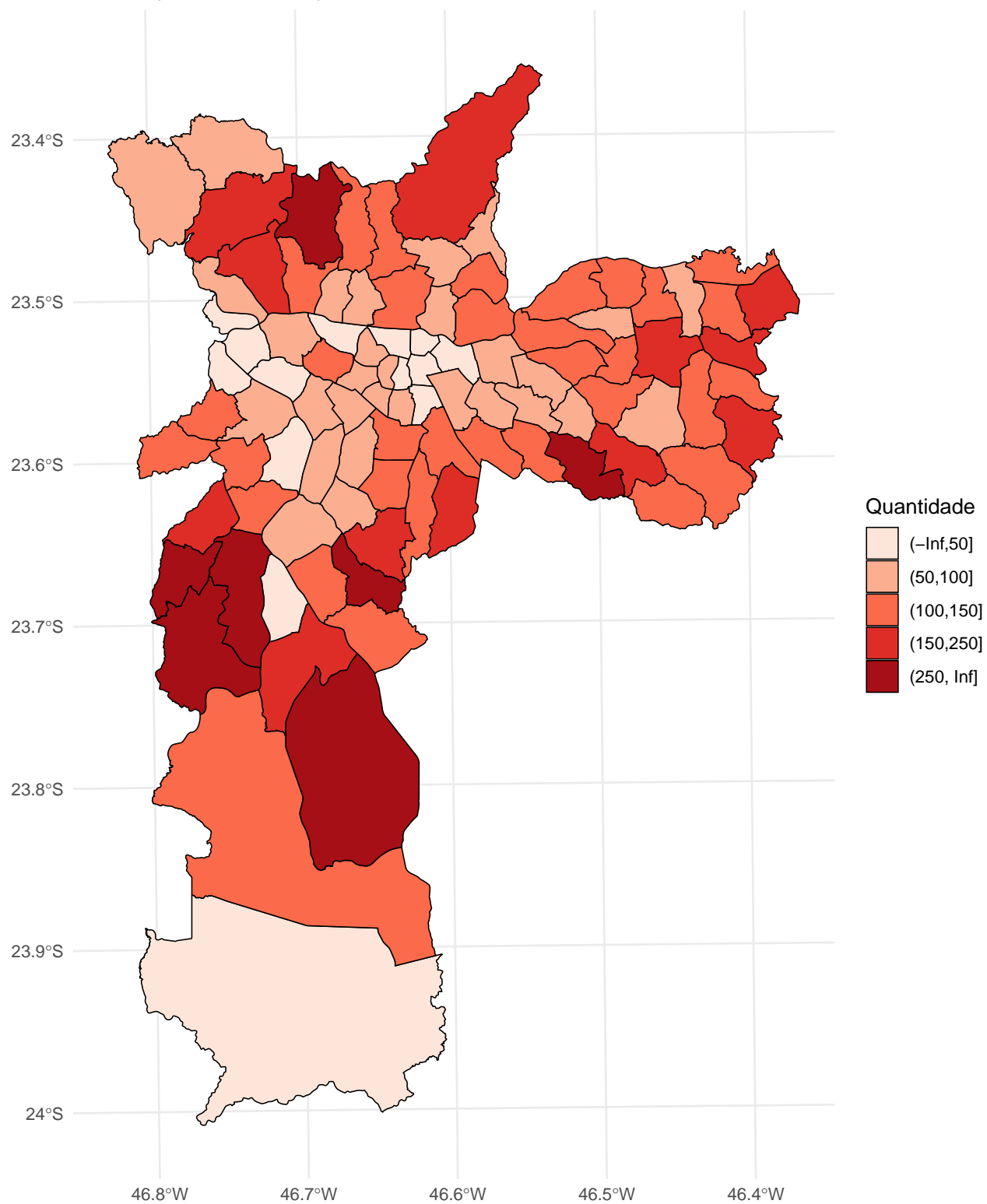
```

cultura <- cultura %>%
  mutate(habitantes_group = cut(habitantes / 1000, breaks = c(-Inf, 50, 100, 150, 250, Inf)))

gPop <- ggplot(cultura, aes(fill=habitantes_group)) +
  geom_sf(color="black", lwd=0.3) +
  scale_fill_brewer(palette="Reds", "Quantidade") +
  theme_minimal() +
  labs(title = "Distribuição da População em São Paulo")
gPop

```

Distibuição da População em São Paulo

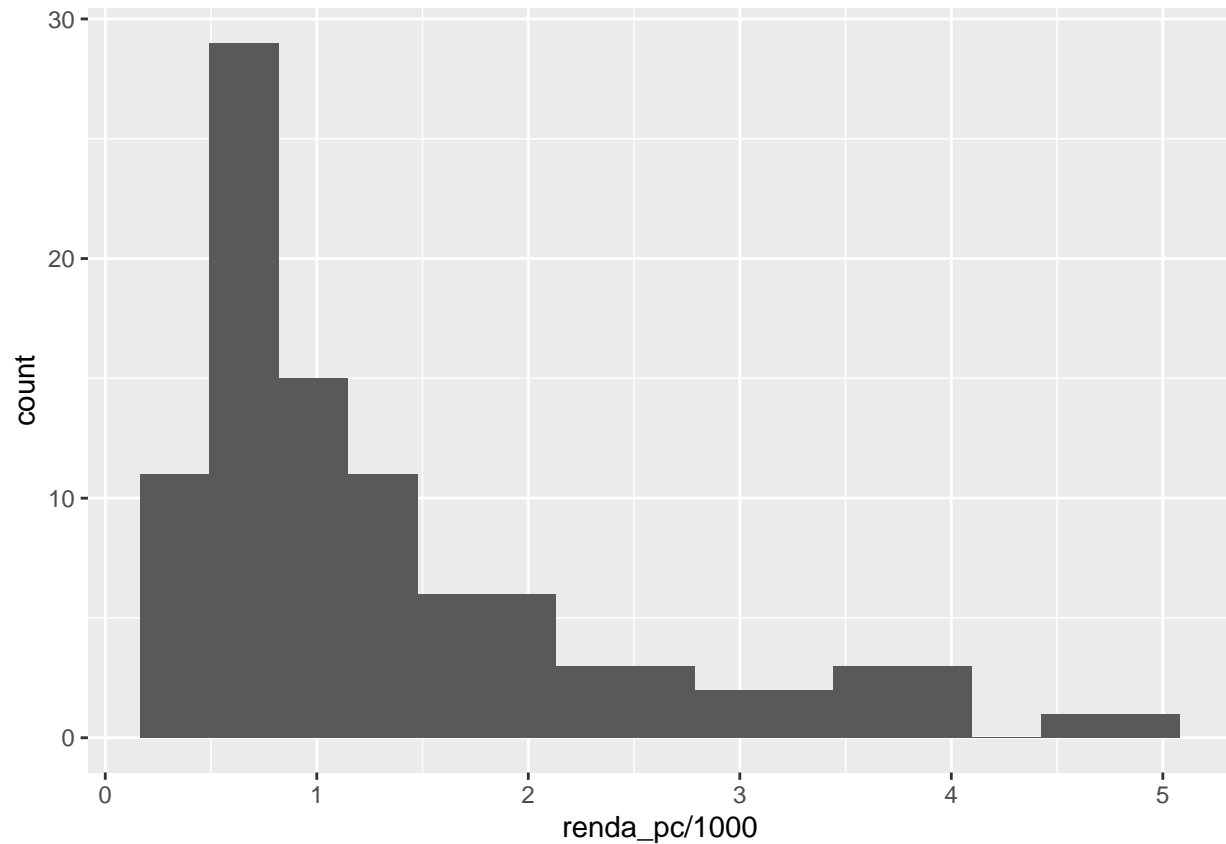


Distribuição da renda

```
cultura %>%
  data.frame %>%
  select(renda_pc) %>%
  summarise_all(funs(min, mean, median, max, sd, IQR))
```

```
##      min    mean  median    max    sd    IQR
## 1 349.8673 1383.59 961.0843 4939.91 1085.404 1175.067
```

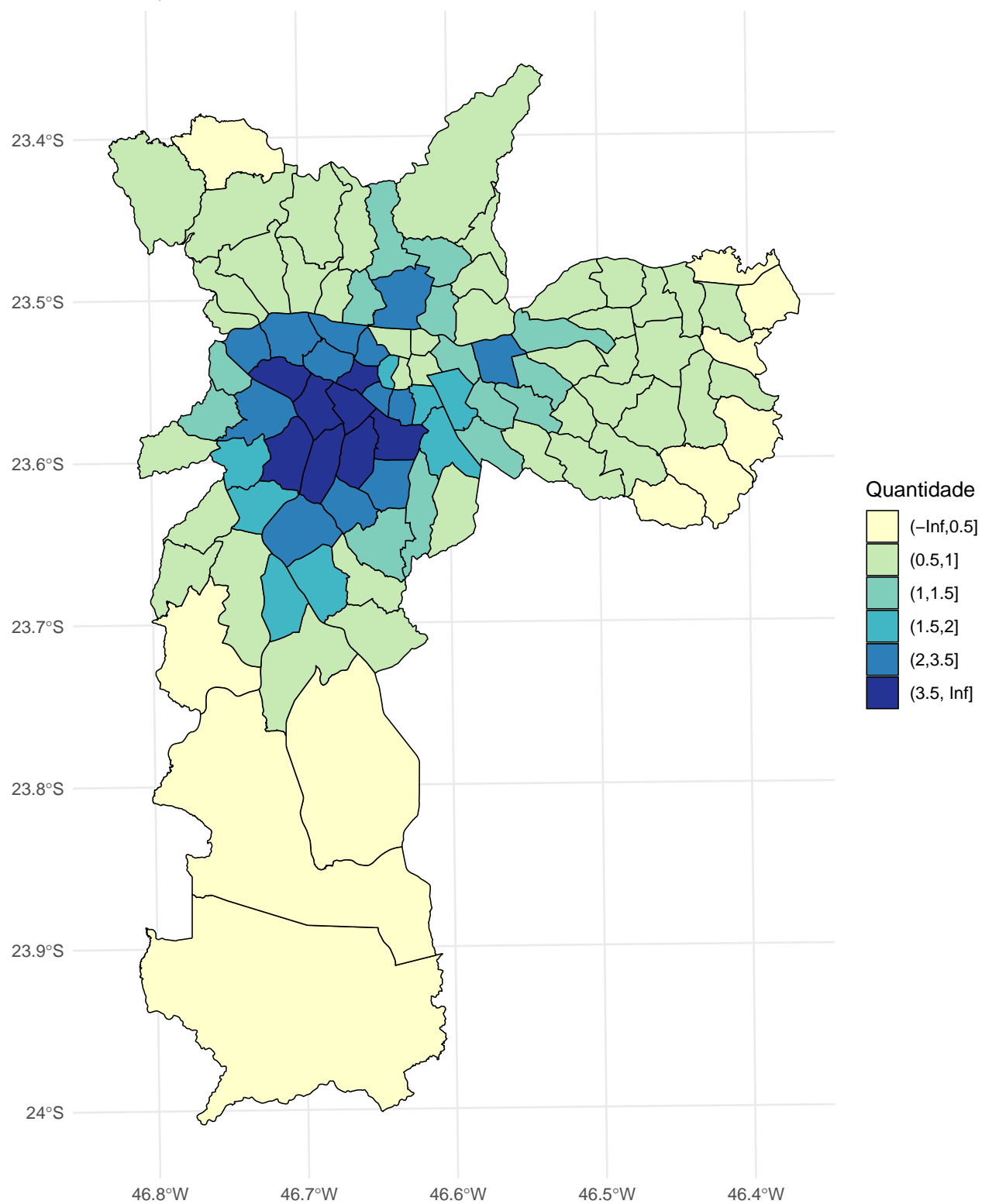
```
ggplot(cultura, aes(x=renda_pc / 1000)) +
  geom_histogram(bins = 15)
```



```
cultura <- cultura %>%
  mutate(renda_pc_group = cut(renda_pc/1000, breaks = c(-Inf, 0.5, 1, 1.5, 2, 3.5, Inf)))

gRenda <- ggplot(cultura, aes(fill=renda_pc_group)) +
  geom_sf(color="black", lwd=0.3) +
  scale_fill_brewer(palette="YlGnBu", "Quantidade") +
  theme_minimal() +
  labs(title = "Distribuição da Renda Per Capita em São Paulo")
gRenda
```

Distibuição da Renda Per Capita em São Paulo

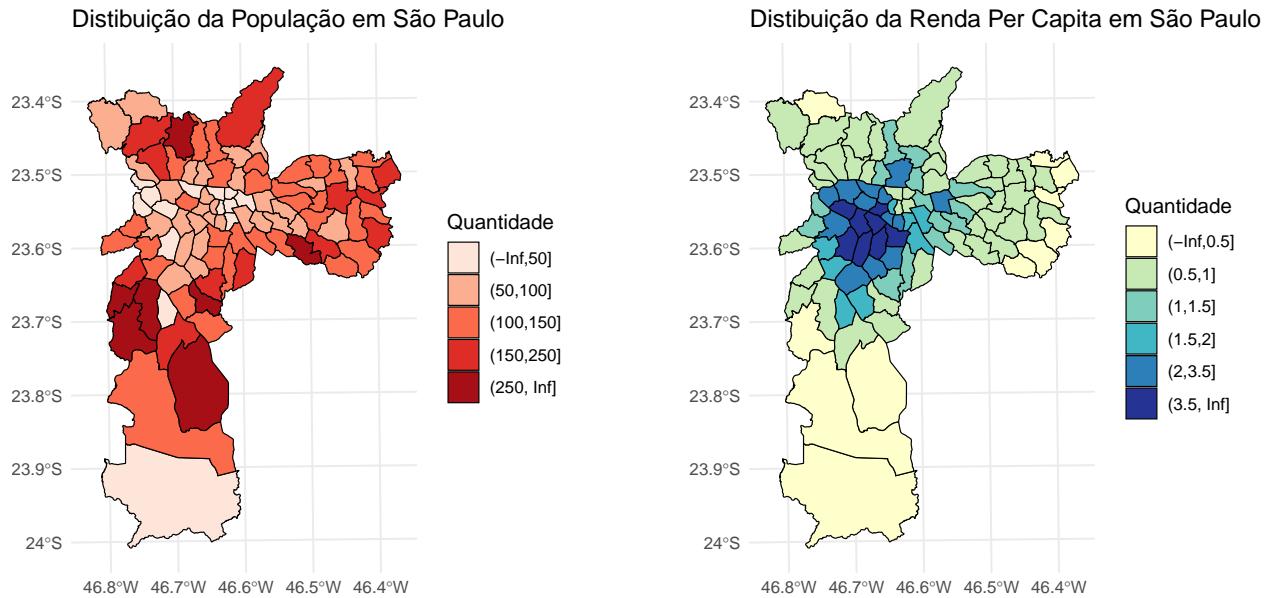


Juntando os dois mapas

```
library(gridExtra)

##
## Attaching package: 'gridExtra'
## The following object is masked from 'package:dplyr':
##
##      combine

grid.arrange(gPop, gRenda, ncol=2)
```



Relação entre as distribuições populacional e de equipamentos

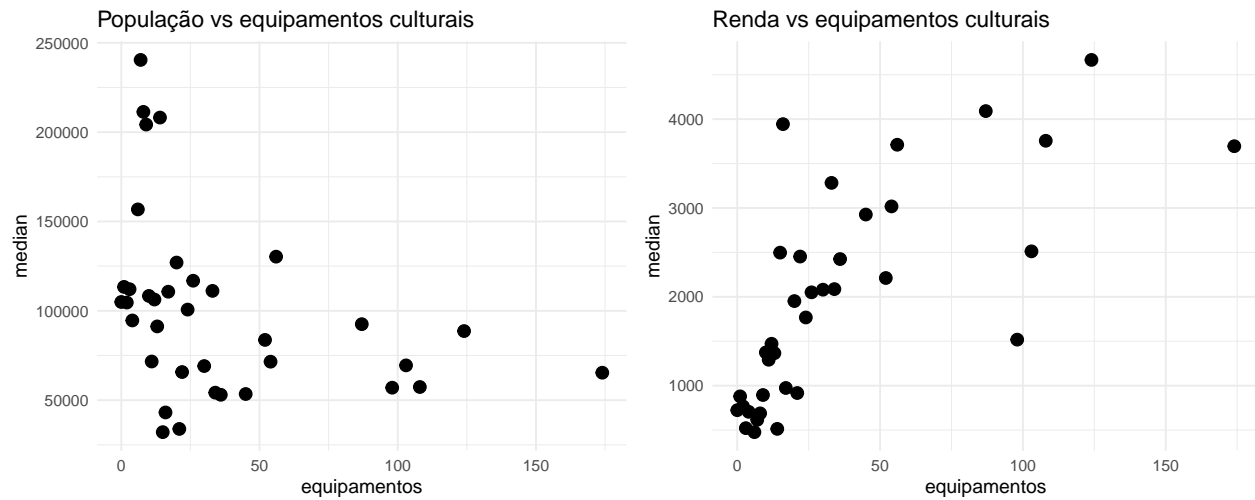
```
gEquipPop <- cultura %>%
  data.frame %>%
  group_by(equipamentos) %>%
  summarise(median = median(habitantes)) %>%
  ggplot(aes(x=equipamentos, y=median)) +
  geom_point(size=3) +
  scale_color_viridis_d("Habitantes (mil)") +
  labs(title = "População vs equipamentos culturais") +
  theme_minimal()

## 'summarise()' ungrouping output (override with '.groups' argument)

gEquipRenda <- cultura %>%
  data.frame %>%
  group_by(equipamentos) %>%
  summarise(median = median(renda_pc)) %>%
  ggplot(aes(x=equipamentos, y=median)) +
  geom_point(size=3) +
  scale_color_viridis_d("Habitantes (mil)") +
  labs(title = "Renda vs equipamentos culturais") +
  theme_minimal()
```

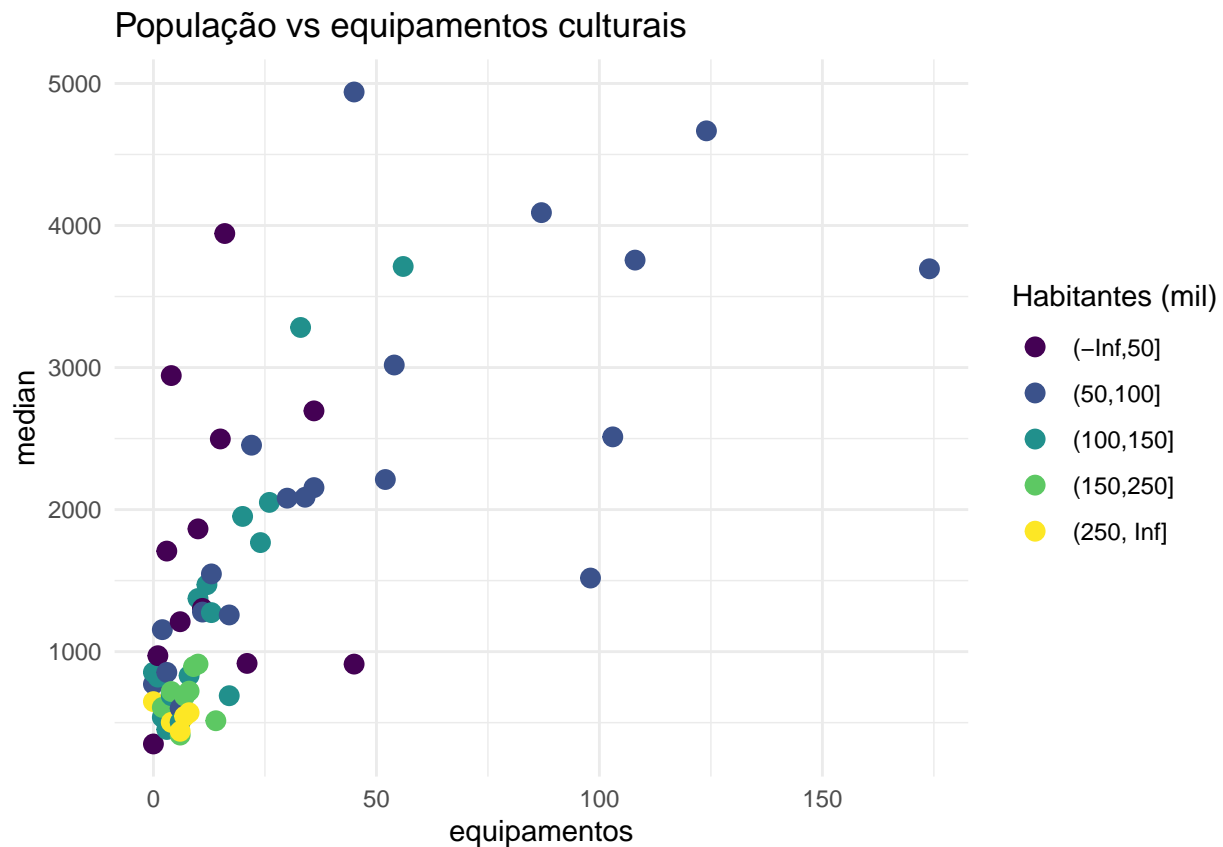
```
## 'summarise()' ungrouping output (override with '.groups' argument)
```

```
grid.arrange(gEquipPop, gEquipRenda, ncol=2)
```



```
cultura %>%  
  data.frame %>%  
  group_by(equipamentos, habitantes_group) %>%  
  summarise(median = median(renda_pc)) %>%  
  ggplot(aes(x=equipamentos, y=median, color=habitantes_group)) +  
  geom_point(size=3) +  
  scale_color_viridis_d("Habitantes (mil)") +  
  labs(title = "População vs equipamentos culturais") +  
  theme_minimal()
```

```
## 'summarise()' regrouping output by 'equipamentos' (override with '.groups' argument)
```



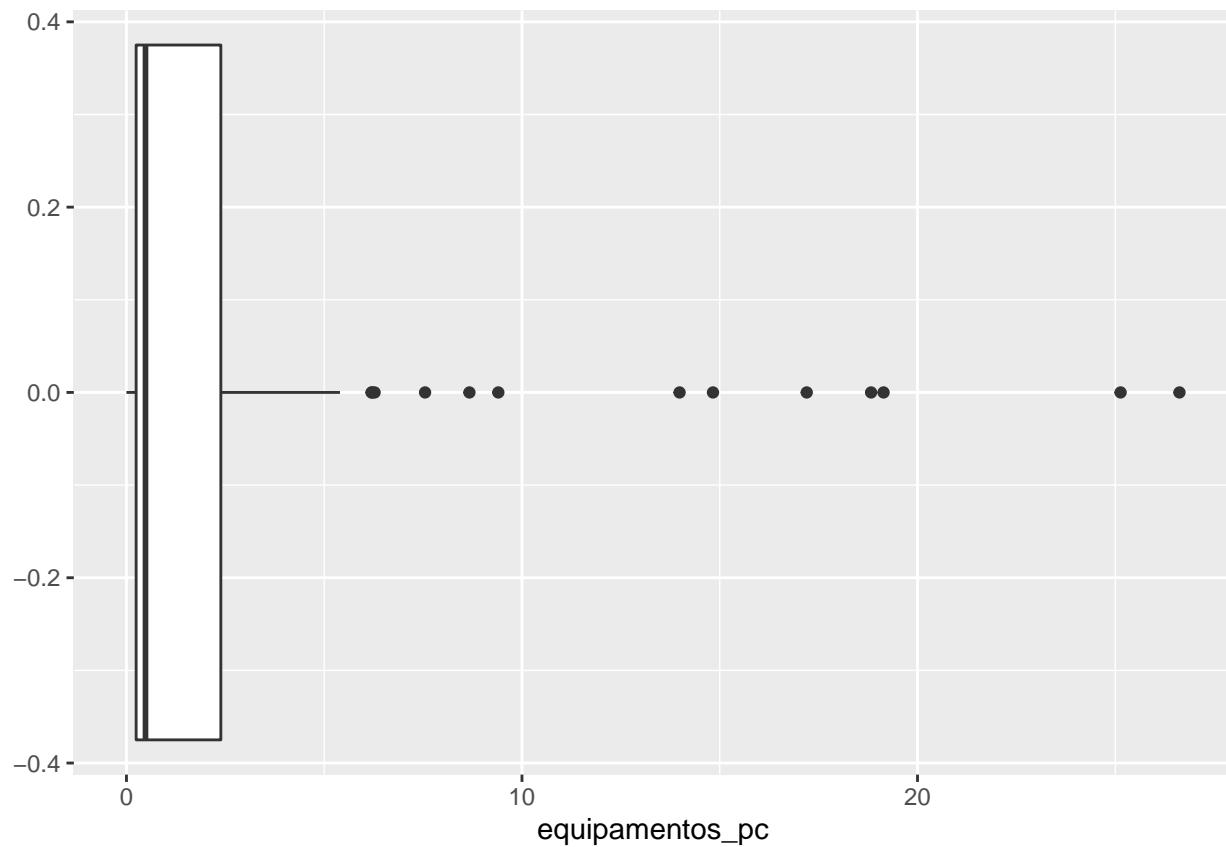
Equipamentos por 10000 habitantes

```
cultura <- cultura %>%
  mutate(equipamentos_pc = (equipamentos / habitantes) * 10000,
         equipamentos_pc_group = cut(equipamentos_pc, breaks = c(-Inf, 0.2, 0.5, 1, 3, 10, Inf)))

summary(cultura$equipamentos_pc)

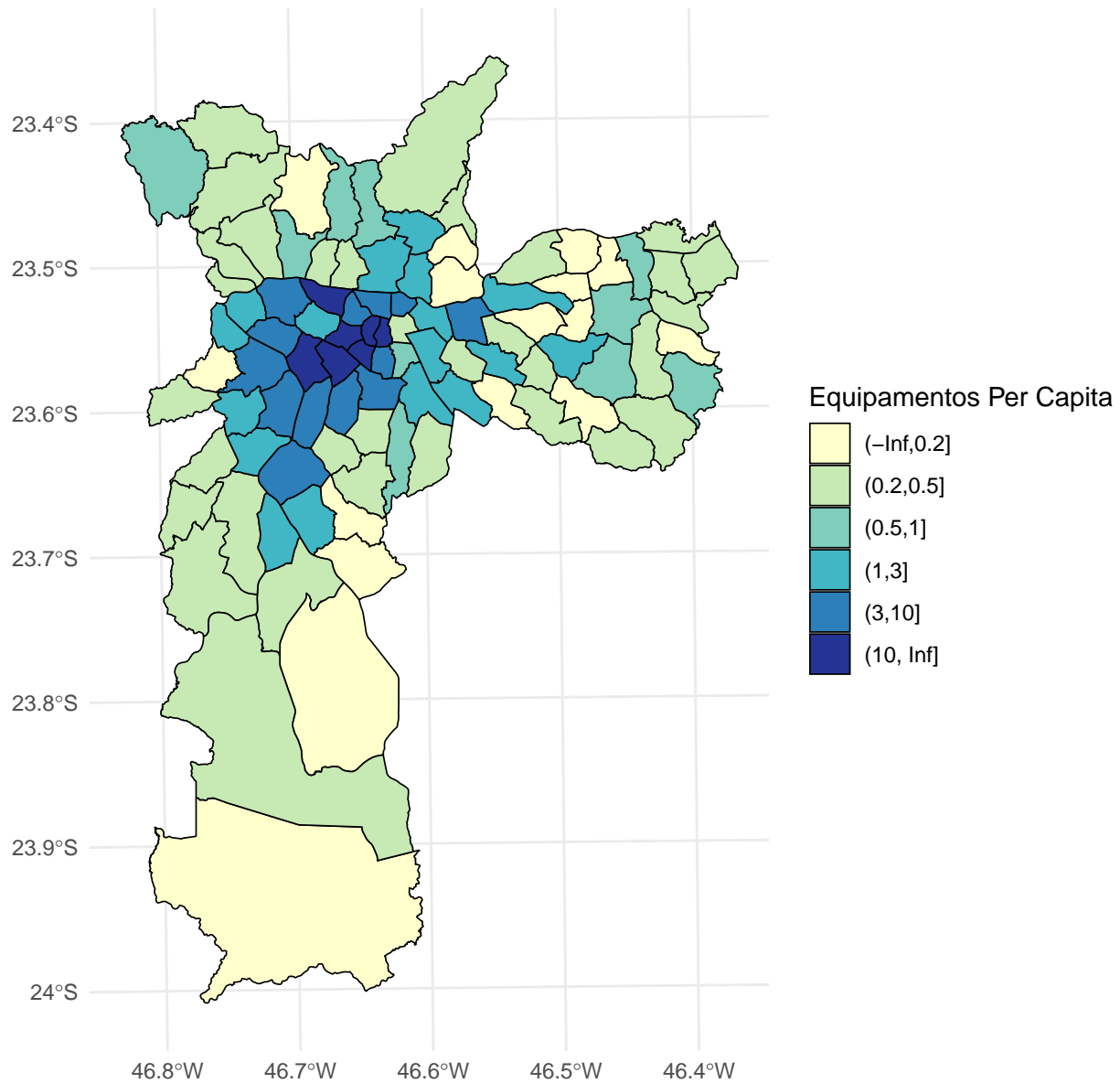
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0000 0.2450 0.4818 2.6762 2.3859 26.6230

ggplot(cultura, aes(x=equipamentos_pc)) +
  geom_boxplot()
```



```
ggplot(cultura, aes(fill=equipamentos_pc_group)) +
  geom_sf(color="black", lwd=0.3) +
  scale_fill_brewer(palette="YlGnBu", "Equipamentos Per Capita") +
  theme_minimal() +
  labs(title = "Distribuição da Equipamentos culturais Per Capita em São Paulo")
```


Distibuição da Equipamentos culturais Per Capita em São Paulo



```
cor(data.frame(cultura) %>% select_if(is.numeric))[, "renda_pc"]
```

```
##      bibliotecas  espacos_culturais      museus      outros
##      -0.08296742    0.56445227    0.48793439    0.04416404
##      teatro_cinema    equipamentos      brancos      pretos
##      0.64481593    0.65752496    -0.17333817    -0.56360087
##      amarelos      pardos      indigenas      domicilios
##      0.28807408    -0.55293433    -0.28130268    -0.28109023
##      renda_domicilios      habitantes      renda_media_dom      renda_pc
##      0.73515735    -0.40859881    0.98644113    1.00000000
##      equipamentos_pc
##      0.54368687
```

```

regLiner <- lm(equipamentos ~ renda_pc + equipamentos_pc, cultura)
summary(regLiner)

##
## Call:
## lm(formula = equipamentos ~ renda_pc + equipamentos_pc, data = cultura)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -79.628  -2.931   0.646   4.553  45.154
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -3.098269   2.298059  -1.348   0.181
## renda_pc       0.007360   0.001546   4.760 7.09e-06 ***
## equipamentos_pc 3.934443   0.320478  12.277 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 13.73 on 93 degrees of freedom
## Multiple R-squared:  0.7834, Adjusted R-squared:  0.7787
## F-statistic: 168.2 on 2 and 93 DF,  p-value: < 2.2e-16

cultura$equipPred <- predict(regLiner, newdata = select(cultura, renda_pc, equipamentos_pc))

ggplot(cultura, aes(x = equipamentos))

```

