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[grepl(".\*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)) / 10^6) %>%   
 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))

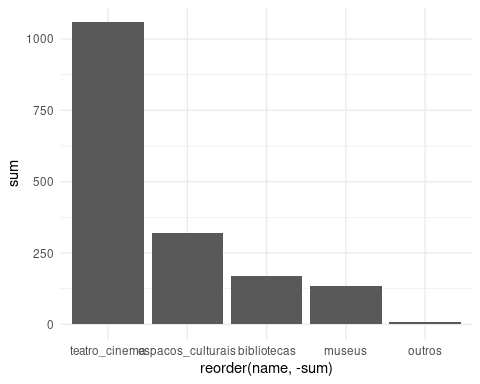
## Warning: `funs()` is deprecated as of dplyr 0.8.0.  
## Please use a list of either functions or lambdas:   
##   
## # Simple named list:   
## list(mean = mean, median = median)  
##   
## # Auto named with `tibble::lst()`:   
## tibble::lst(mean, median)  
##   
## # Using lambdas  
## list(~ mean(., 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

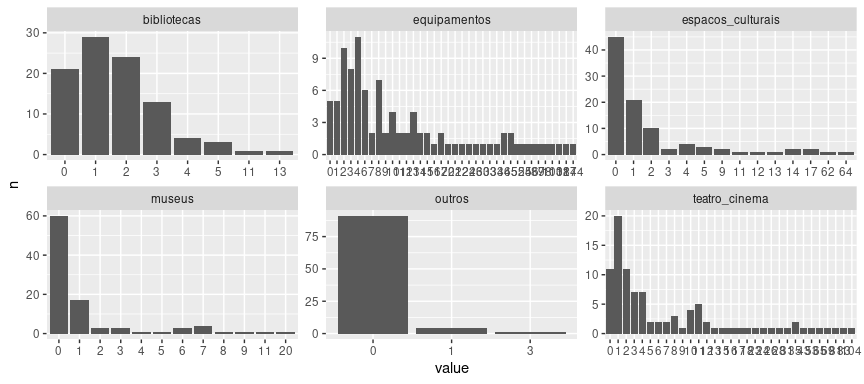
### Quantidades por tipo de equipamento

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")



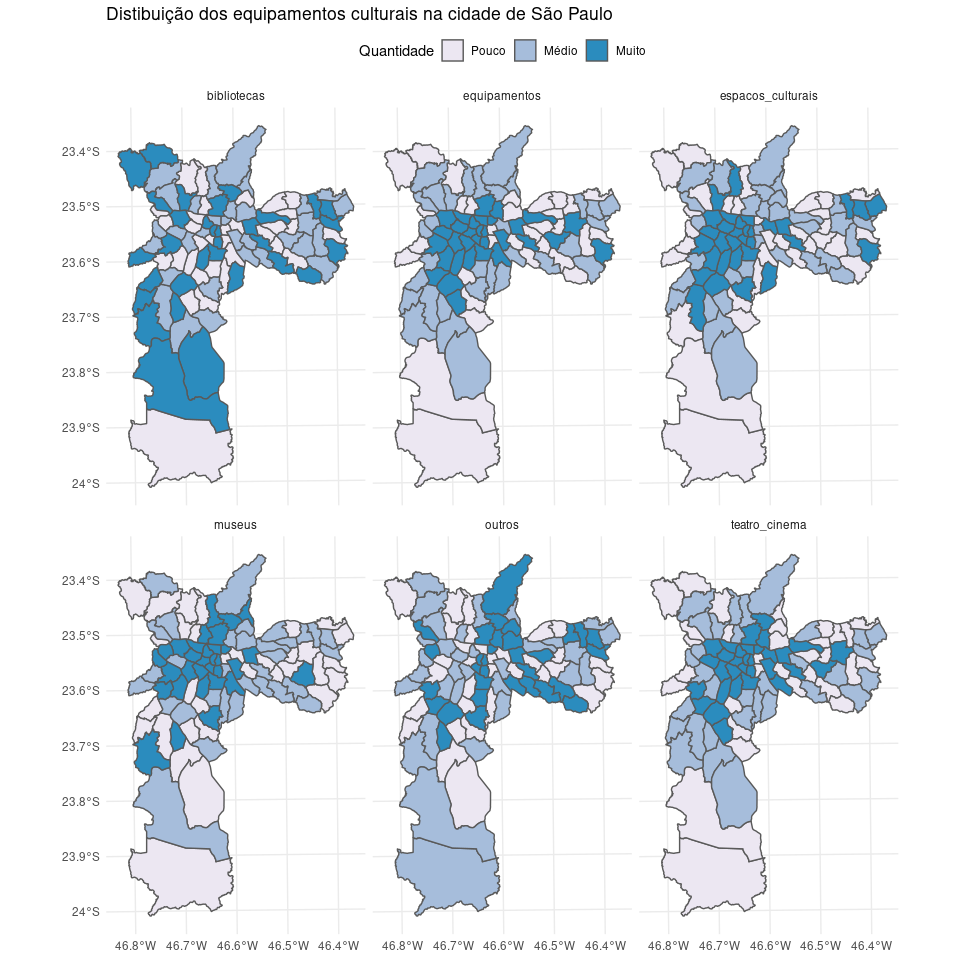
### Distribuição territorial dos equipamentos

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 = "Distibuição 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

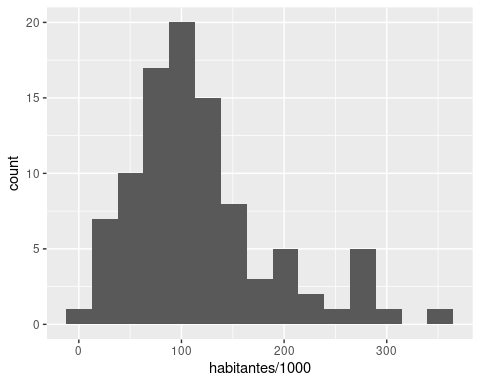


### Distribuição da população

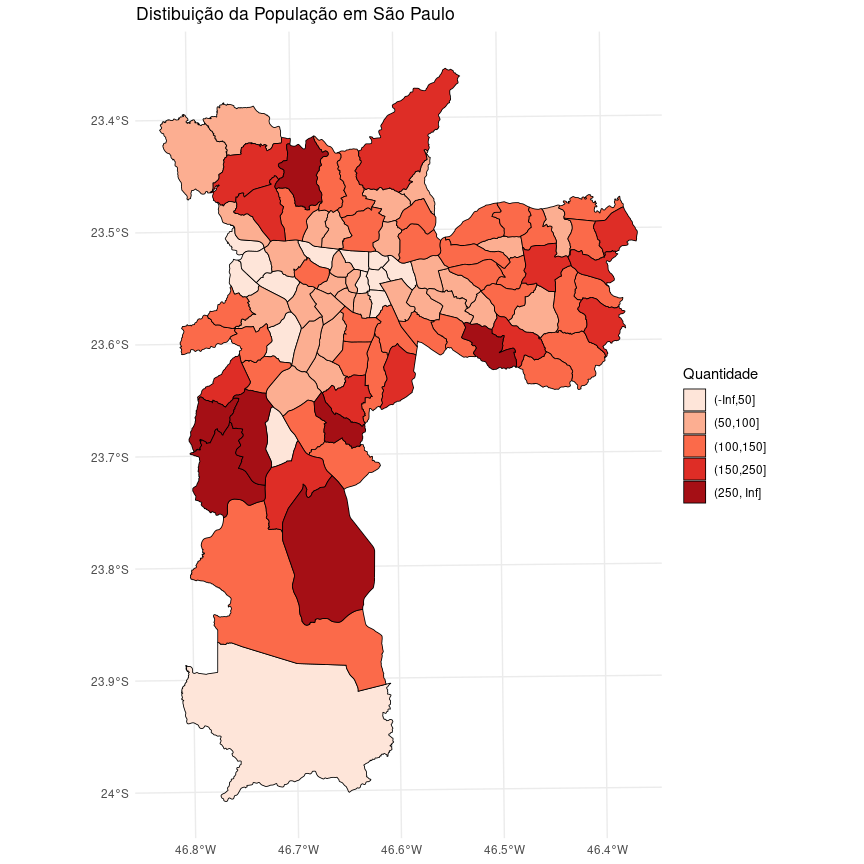
censo <- read.csv("../dados/censo.csv") %>%   
 mutate(habitantes = indigenas + pretos + pardos + brancos + amarelos,  
 renda\_media\_dom = renda\_domicilios / domicilios,  
 renda\_pc = renda\_domicilios / habitantes)  
  
cultura <- cultura %>%   
 merge(censo, by = "cod\_distrito")  
  
glimpse(cultura)

## 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, 1,…  
## $ museus <dbl> 0, 1, 0, 0, 0, 2, 6, 0, 7, 0, 0, 20, 0, 1, 0, 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, 23…  
## $ equipamentos <dbl> 2, 16, 4, 2, 2, 36, 103, 11, 21, 1, 4, 34, 8, 3, 2,…  
## $ nome\_distrito <chr> "ÁGUA RASA", "ALTO DE PINHEIROS", "ANHANGUERA", "AR…  
## $ brancos <int> 71095, 38262, 32321, 62365, 64722, 11640, 51075, 32…  
## $ pretos <int> 2154, 860, 4282, 4914, 10406, 438, 3377, 1777, 1552…  
## $ amarelos <int> 2087, 1361, 333, 2532, 1452, 417, 3253, 1192, 5141,…  
## $ pardos <int> 9565, 2615, 28783, 19762, 28622, 1811, 11659, 9354,…  
## $ indigenas <int> 62, 19, 70, 49, 67, 17, 96, 189, 138, 123, 261, 51,…  
## $ domicilios <int> 28652, 15408, 19437, 27661, 33712, 5623, 29967, 143…  
## $ renda\_domicilios <int> 114911602, 170039733, 34768395, 71198307, 77883540,…  
## $ habitantes <int> 84963, 43117, 65789, 89622, 105269, 14323, 69460, 4…  
## $ 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, 739.8526,…  
## $ geometry <MULTIPOLYGON [m]> MULTIPOLYGON (((338561.8 73..., MULTIP…

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 = "Distibuição da População em São Paulo")  
gPop

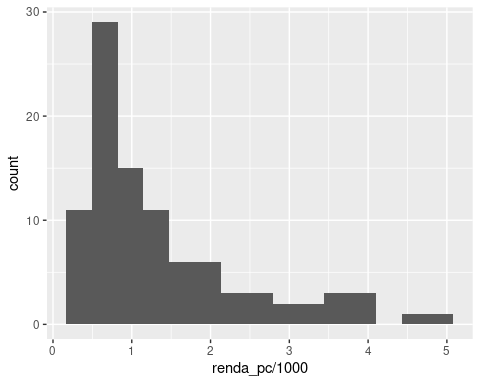


### Distribuição da renda

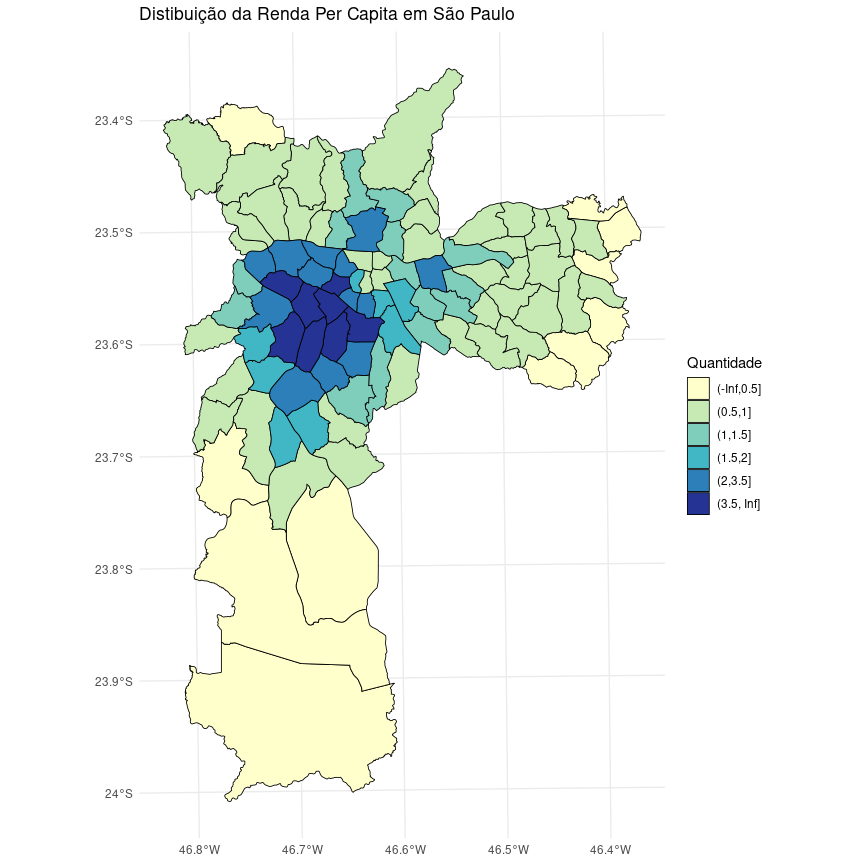
cultura %>%   
 data.frame %>%   
 select(renda\_pc) %>%   
 summarise\_all(funs(min, mean, median, max, sd, IQR))

## 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 = "Distibuição da Renda Per Capita em São Paulo")  
gRenda



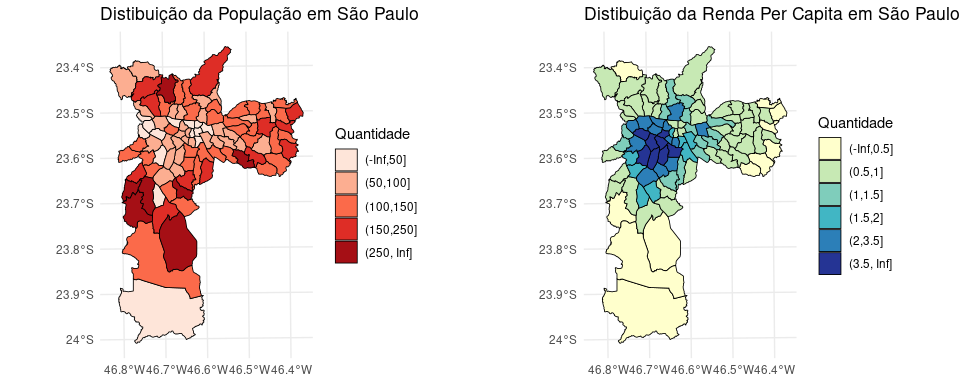
### Juntando os dois mapas

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

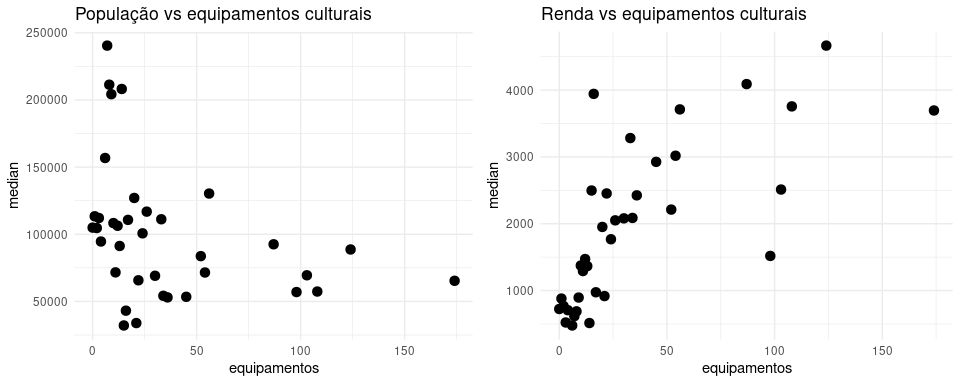
gEquipPop <- cultura %>%   
 data.frame %>%   
 group\_by(equipamentos) %>%   
 summarise(median = median(habitantes)) %>%   
 ggplot(aes(x=equipamentos, y=median)) +  
 geom\_point(size=3) +  
 scale\_color\_viridis\_d("Habitantes (mil)") +  
 labs(title = "População vs equipamentos culturais") +  
 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()

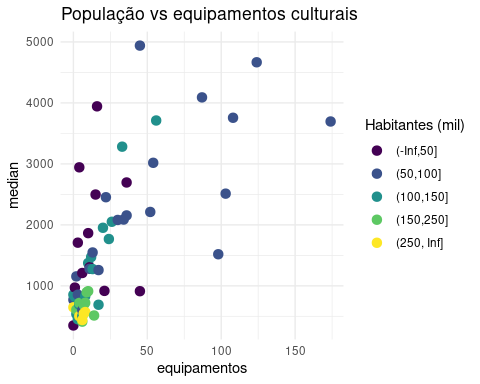
## `summarise()` ungrouping output (override with `.groups` argument)

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)



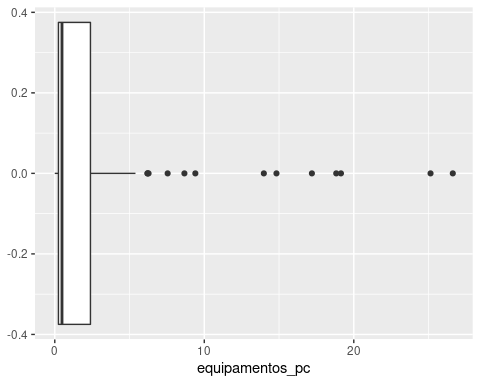
## Equipamentos por 10000 habitantes

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)))

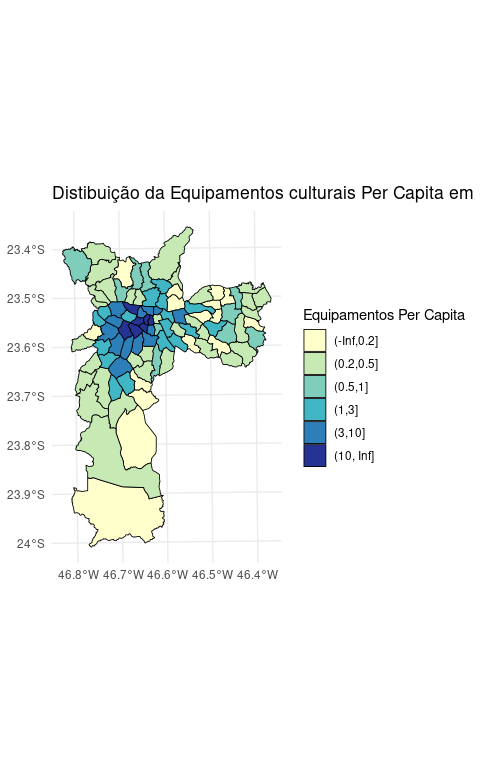
summary(cultura$equipamentos\_pc)

## 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 = "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))

