

# Planea 2019

First we connect to our Postgres database

And we load the PLANEA 2019 results

Change results type to numeric (TODO: set this schema in the airflow ingest pipeline)

We need to change some municipalities names in order to join this database with the CONAPO one. This are the missed-matched names in the tables of the database. I'll change the ones from the schools.

Load the CONAPO table

Load the marginalization table and change the variable names to usable ones.

```
## [1] TRUE
```

Join the three tables. I should run this in Postgres. *TODO* write SQL statement. But for this we should have the same values from ingest.

gm	Number of municipalities
Very Low	690
Low	996
Medium	1028
High	1634
Very High	566

How many schools by Marginalization Level

gm	Number of schools
Very Low	52612
Low	24796
Medium	22832
High	27108
Very High	9488

How many schools by School Type

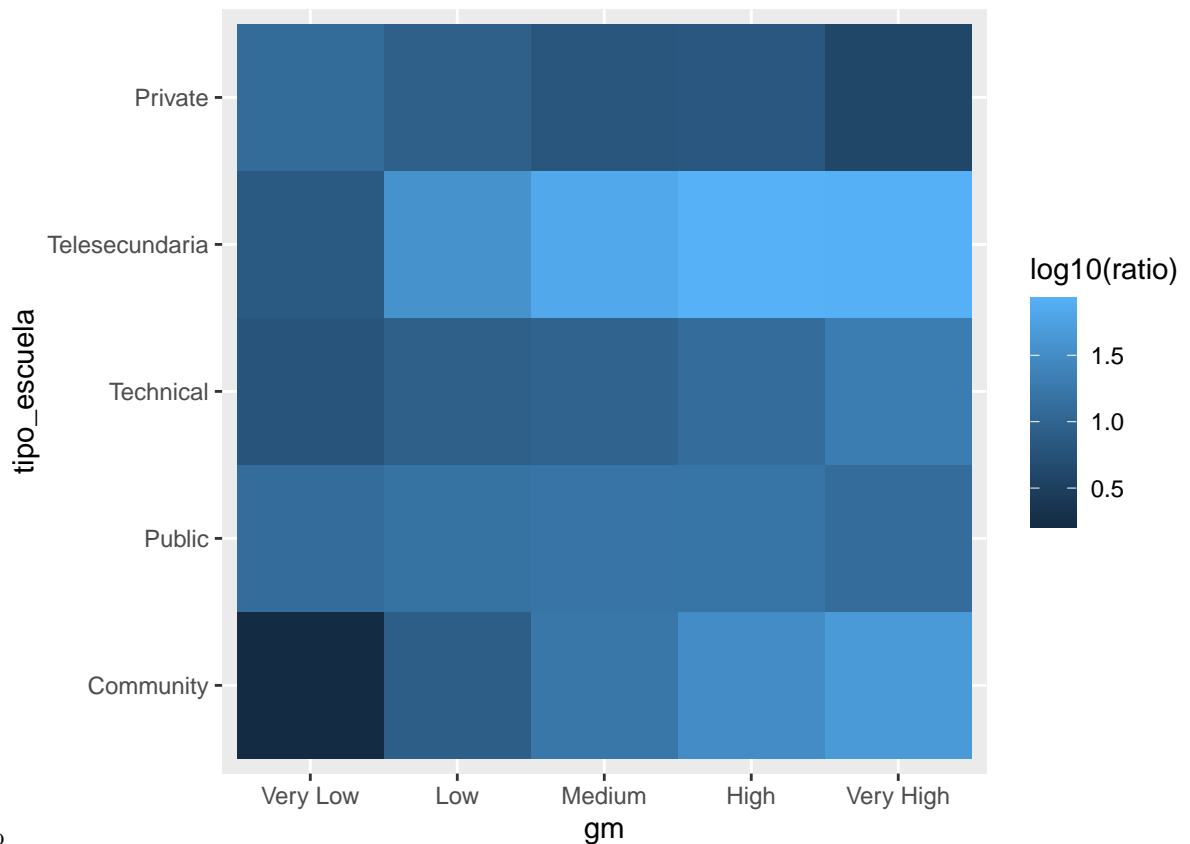
tipo_escuela	Number of schools
Community	13736
Public	27364
Technical	14804
Telesecundaria	60272
Private	20660

Number of schools by type and marginalization level

gm	Community	Public	Technical	Telesecundaria	Private
Very Low	888	17528	8424	9100	16672
Low	1452	4812	2752	13120	2660
Medium	2880	2796	1528	14812	816
High	5420	1932	1480	17828	448
Very High	3096	296	620	5412	64

Schools to population ratio by type and marginalization level

gm	Community	Public	Technical	Telesecundaria	Private
Very Low	1.601999	12.47991	6.073040	7.051799	11.901271
Low	8.138358	14.58139	8.672817	36.413071	8.586183
Medium	17.172003	15.51260	9.699112	66.439866	6.485102
High	31.201097	16.16145	12.525018	85.842465	6.650130
Very High	46.011449	12.67822	20.024650	83.566234	3.990174

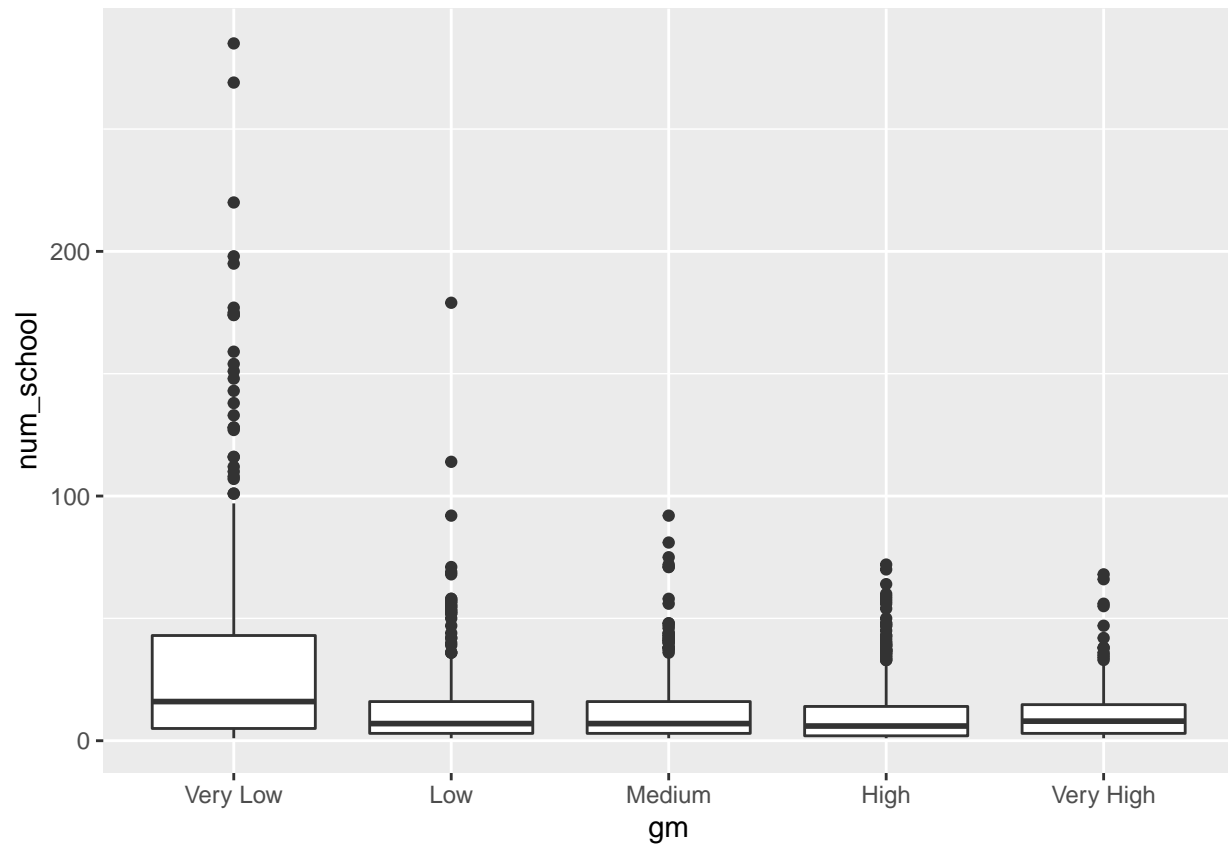


Heatmap of the ratio

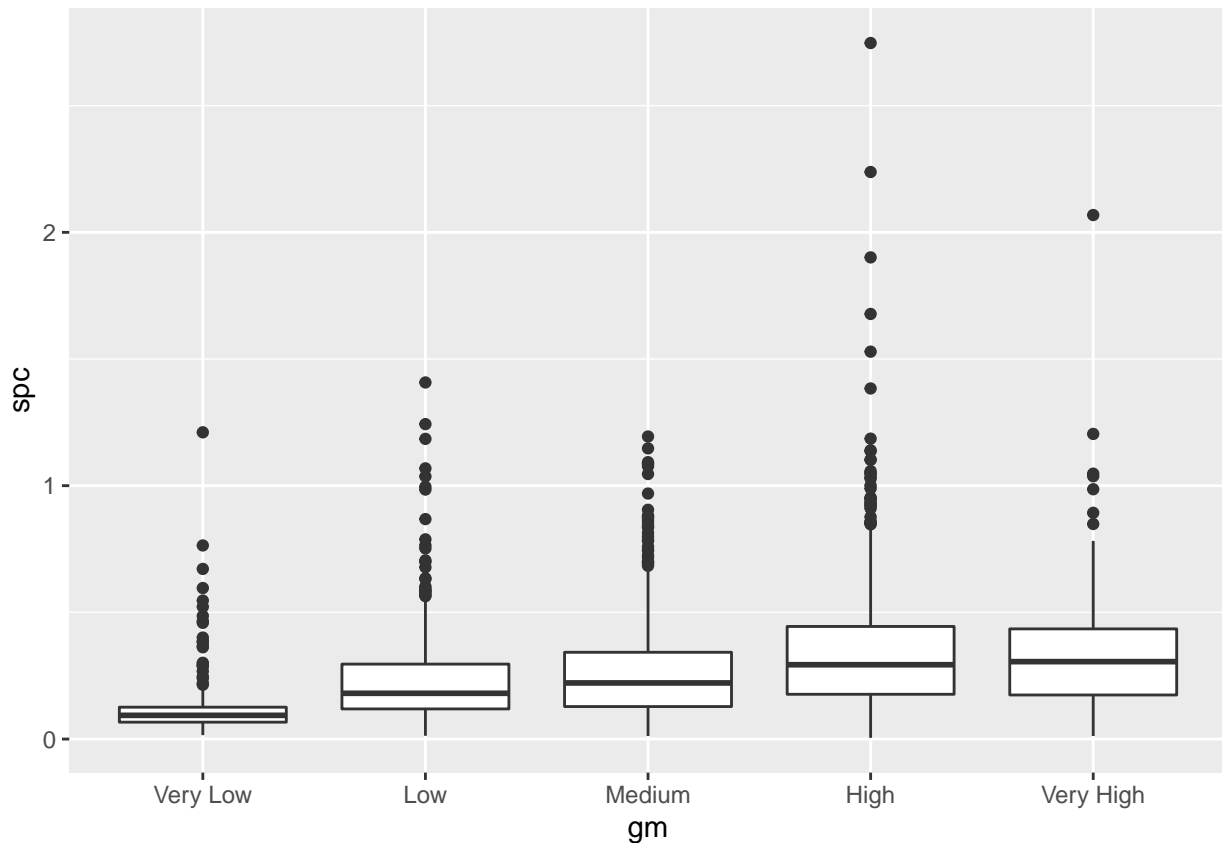
Select only the schools with more than 50% of the students tested at both Language and Math

```
schools <- schools[(schools$porc_len > 50 & schools$porc_mat > 50),]
mun_schools <- schools %>%
  select(CLAVE,escuela,POP,marginacion,gm,elementary,illiteracy) %>%
  unique() %>% dplyr::group_by(CLAVE) %>%
  dplyr::summarise(num_school = n(),POB=unique(POP),elementary=unique(elementary),illiteracy=unique(illiteracy))
  mutate(spc = num_school/(POB/1000))
```

```
ggplot(mun_schools)+
  geom_boxplot(aes(gm, num_school))
```



```
ggplot(mun_schools)+
  geom_boxplot(aes(gm, spc))
```



```
mar_scores <- schools %>% select(CLAVE,POP,illiteracy,elementary,I_porcen,II_porcen,III_porcen,IV_porcen,I_porcen_mat,II_porcen_mat,III_porcen_mat,IV_porcen_mat)

schools_l <- schools %>% group_by(CLAVE, clave, escuela, gm) %>% summarise(num_school=n(),POP=unique(POP),illiteracy=unique(illiteracy),elementary=unique(elementary),I_porcen=unique(I_porcen),II_porcen=unique(II_porcen),III_porcen=unique(III_porcen),IV_porcen=unique(IV_porcen),I_porcen_mat=median(I_porcen_mat),II_porcen_mat=median(II_porcen_mat),III_porcen_mat=median(III_porcen_mat),IV_porcen_mat=median(IV_porcen_mat))

mar_scores %<>% group_by(CLAVE) %>% dplyr::summarise(num_school=n(),POP=unique(POP),illiteracy=unique(illiteracy),elementary=unique(elementary),I_porcen=unique(I_porcen),II_porcen=unique(II_porcen),III_porcen=unique(III_porcen),IV_porcen=unique(IV_porcen),I_porcen_mat=median(I_porcen_mat),II_porcen_mat=median(II_porcen_mat),III_porcen_mat=median(III_porcen_mat),IV_porcen_mat=median(IV_porcen_mat))
```

Translate some variables to english. There are some with peculiar name that shouldn't be translated.

```
schoools_tidy <- schools %>% select(tipo_escuela,gm,I_porcen,II_porcen,III_porcen,IV_porcen,I_porcen_mat,II_porcen_mat,III_porcen_mat,IV_porcen_mat)
schools_tidy %<>% pivot_longer(c(I_porcen,II_porcen,III_porcen,IV_porcen,I_porcen_mat,II_porcen_mat,III_porcen_mat,IV_porcen_mat),names_to="variable",values_to="value")

results <- c("Language I", "Language II", "Language III", "Language IV", "Math I", "Math II", "Math III", "Math IV")
resultados <- c("I_porcen","II_porcen","III_porcen","IV_porcen","I_porcen_mat","II_porcen_mat","III_porcen_mat","IV_porcen_mat")

schools_tidy$name <- plyr::mapvalues(schools_tidy$name, to=results, from=resultados)

# schools_tidy$tipo_escuela <- factor(schools_tidy$tipo_escuela,levels=c("General Pública","Técnica Pública"))
```

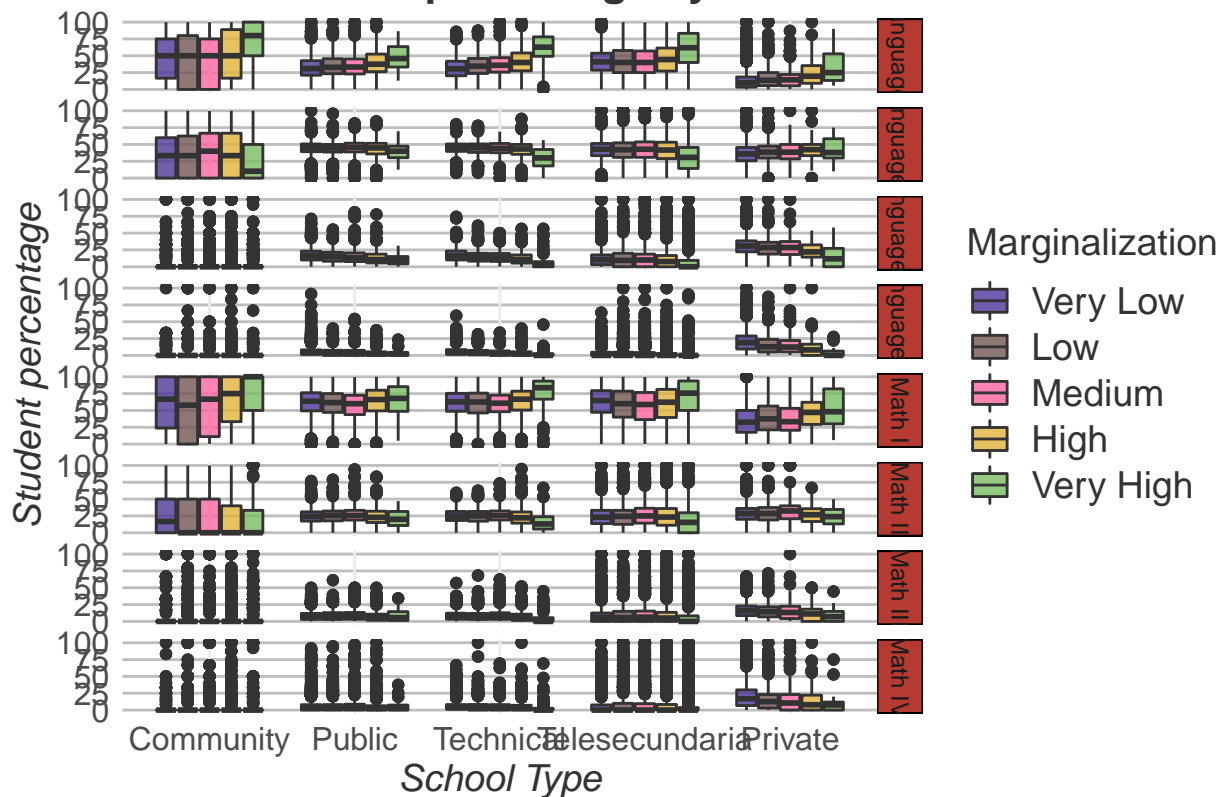
Define the colors

```
altoc <- "#e3b23c"
bajoc <- "#725752"
medioc <- "#fe64a3"
maltoc <- "#78bc61"
mbajoc <- "#4f359b"
```

```
mt <- ggplot(schools_tidy, fig.width=48, fig.height=40)+
  geom_boxplot(aes(tipo_escuela,value,fill=gm),alpha=0.8)+
  scale_fill_manual(name = "Marginalization", values = c("High" = altoc, "Low" = bajoc, "Medium" = medio, "Very Low" = muybajoc, "Very High" = muyaltoc),
    labels = c("Very Low", "Low", "Medium", "High", "Very High"))+
  scale_x_discrete(name="School Type")+
  scale_y_continuous(name = "Student percentage")+
  # theme_classic()+
  ggtitle("School-wise student percentage by achievement level")+
  th+
  theme(
    axis.text.x = element_text(angle = 0),
    panel.grid.minor = element_blank())

mt + facet_grid(vars(name))+
  theme(strip.background = element_rect(fill="#b53a31"))
```

### School-wise student percentage by achievement level



```
comc <- "#efecca"
privc <- "#db2b39"
pubgc <- "#f7ff58"
pubtc <- "#ff934f"
telec <- "#29335c"

mt <- ggplot(schools_tidy)+
  geom_boxplot(aes(gm,value,fill=tipo_escuela),alpha=0.8)+
  scale_fill_manual(name = "School Type", values = c("Community" = comc, "Private" = privc, "Public" = pubgc, "Technical" = pubtc, "Telesecundaria" = telec))+
  scale_x_discrete(name="Marginalization")+
  scale_y_continuous(name = "Student percentage")+
  # theme_classic()+
  ggtitle("Student percentage by school type and marginalization level")+
  th+
  theme(
    axis.text.x = element_text(angle = 0),
    panel.grid.minor = element_blank())

mt + facet_grid(vars(name))+
  theme(strip.background = element_rect(fill="#b53a31"))
```

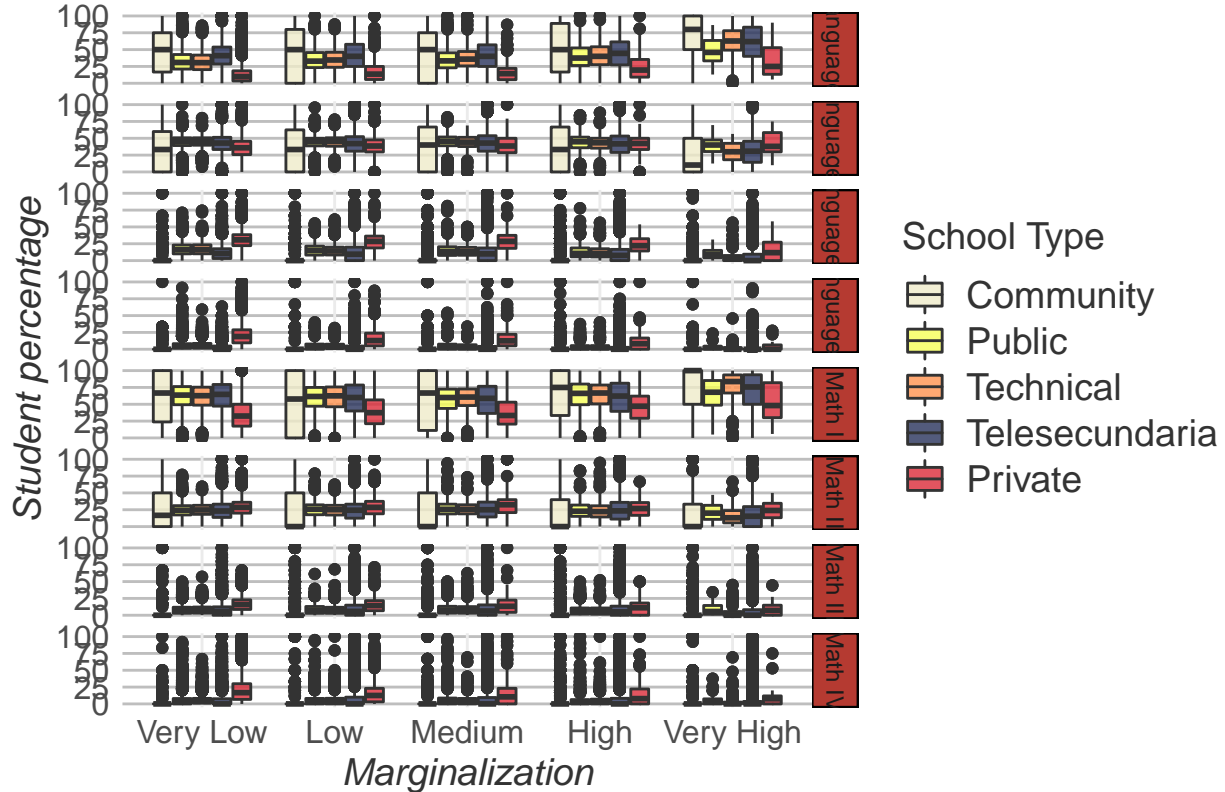
```

ggtitle("School-wise student percentage by achievement level")+
th+
theme(
  axis.text.x = element_text(angle = 0),
  panel.grid.minor = element_blank())

mt + facet_grid(vars(name)) +
  theme(strip.background = element_rect(fill="#b53a31"))

```

## chool-wise student percentage by achievement level



```
require(gridExtra)
```

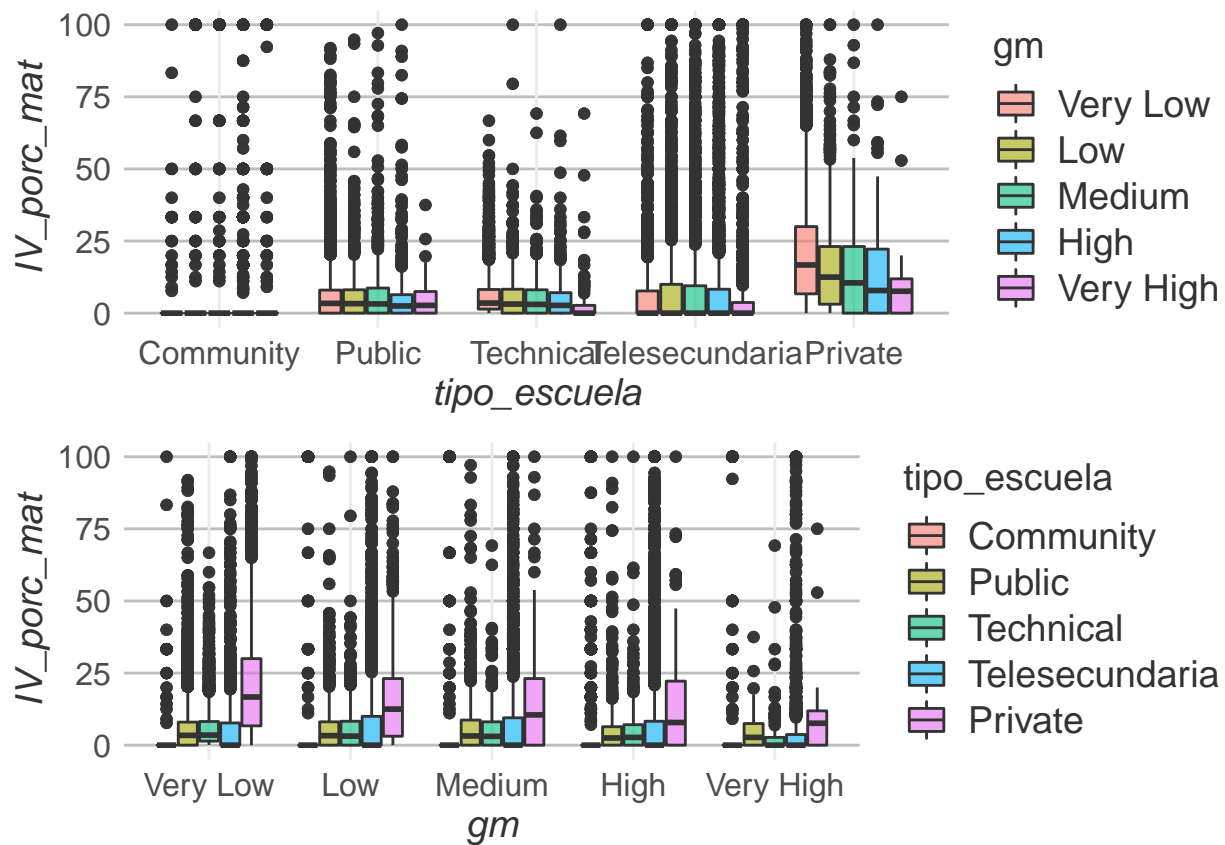
```

uno <- ggplot(schools)+
  geom_boxplot(aes(tipo_escuela,IV_porcentaje_mat,fill=gm),alpha=0.6)+
  th+
  theme(
    axis.text.x = element_text(angle = 0),
    panel.grid.minor = element_blank())

dos <- ggplot(schools)+
  geom_boxplot(aes(gm,IV_porcentaje_mat,fill=tipo_escuela),alpha=0.6)+
  th+
  theme(
    axis.text.x = element_text(angle = 0),
    panel.grid.minor = element_blank())

grid.arrange(uno, dos, ncol=1)

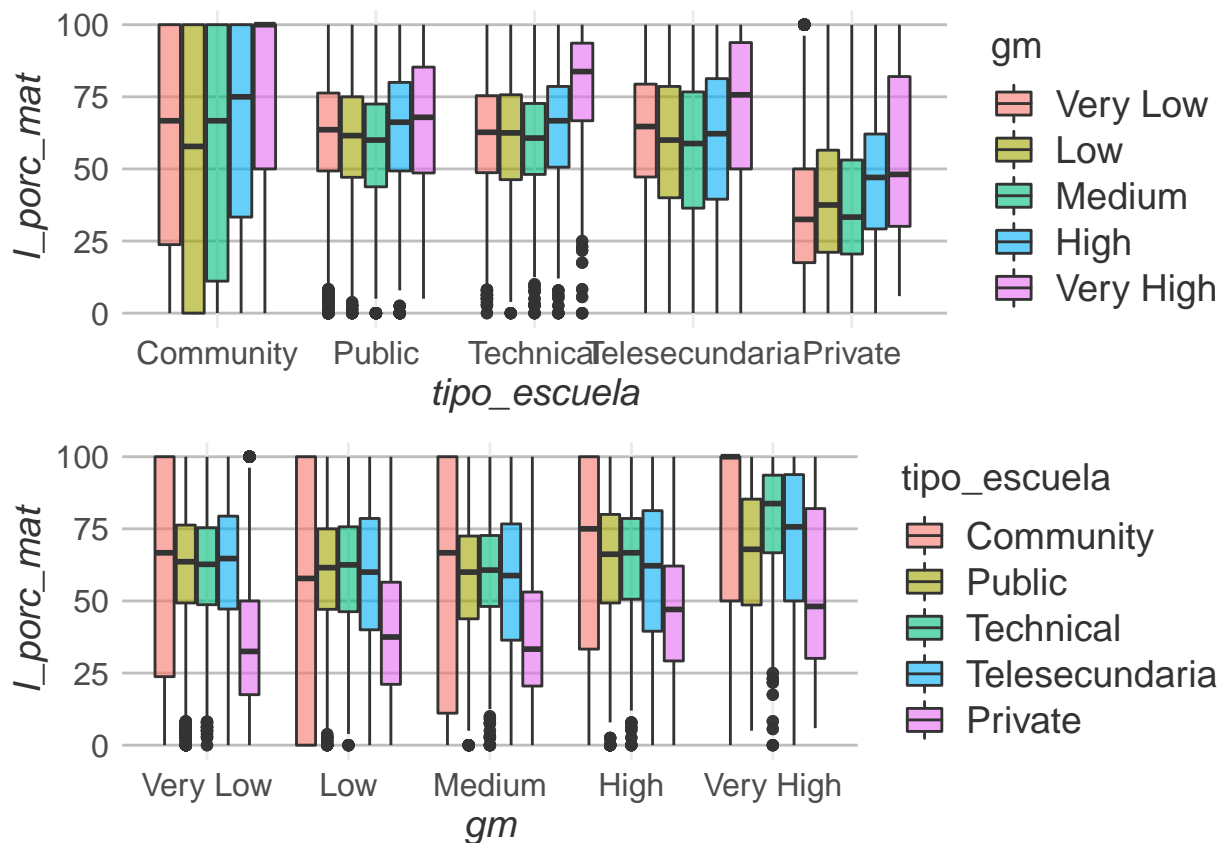
```



```
uno <- ggplot(schools)+
  geom_boxplot(aes(tipo_escuela,I_porc_mat,fill=gm),alpha=0.6)+
  th+
  theme(
    axis.text.x = element_text(angle = 0),
    panel.grid.minor = element_blank())

dos <- ggplot(schools)+
  geom_boxplot(aes(gm,I_porc_mat,fill=tipo_escuela),alpha=0.6)+
  th+
  theme(
    axis.text.x = element_text(angle = 0),
    panel.grid.minor = element_blank())

grid.arrange(unos, dos, ncol=1)
```

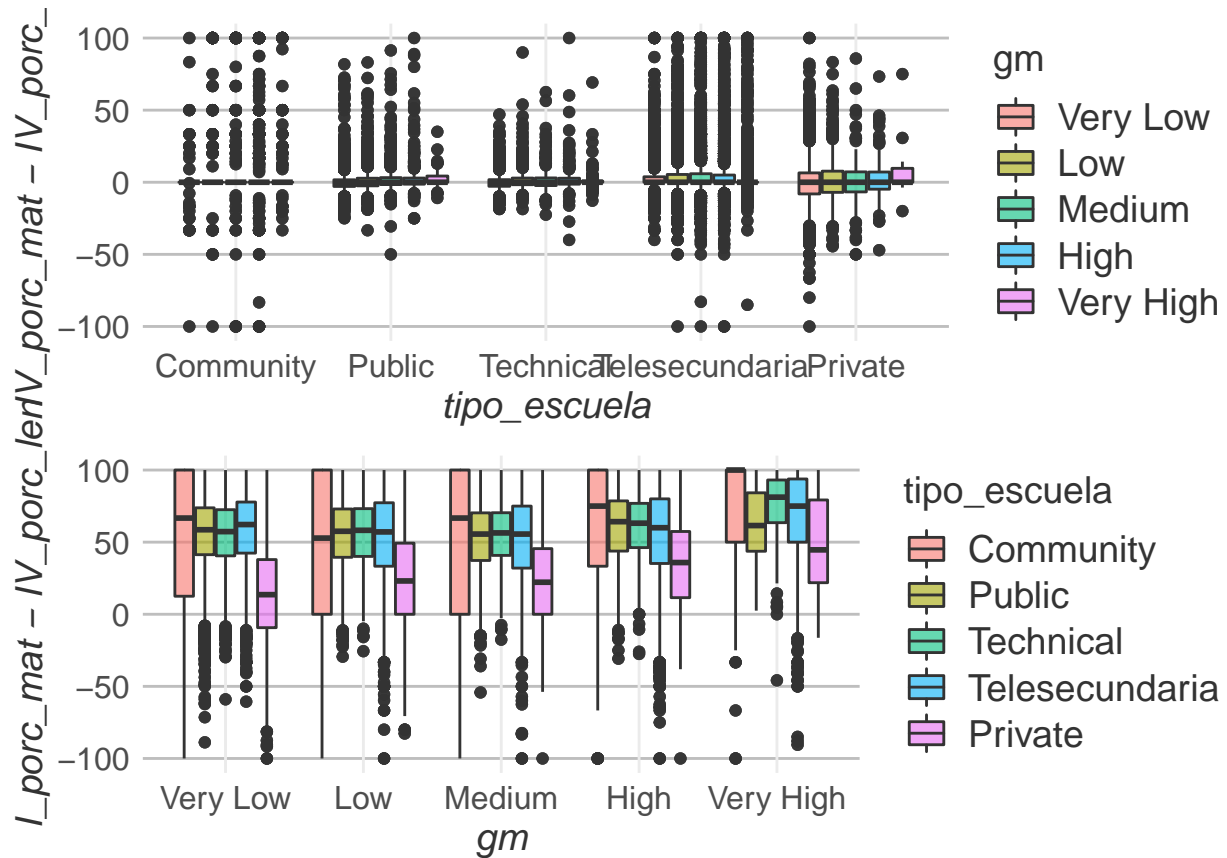


```
uno <- ggplot(schools)+
  geom_boxplot(aes(tipo_escuela,IV_porc_mat-IV_porc_len,fill=gm),alpha=0.6)+
  th+
  theme(
    axis.text.x = element_text(angle = 0),
    panel.grid.minor = element_blank())

dos <- ggplot(schools)+
  geom_boxplot(aes(gm,I_porc_mat-IV_porc_len,fill=tipo_escuela),alpha=0.6)+
  th+
  theme(
    axis.text.x = element_text(angle = 0),
    panel.grid.minor = element_blank())

grid.arrange(unos, dos, ncol=1)
```





```
uno <- ggplot(schools)+
  geom_boxplot(aes(tipo_escuela,I_porc_mat-I_porc_len,fill=gm),alpha=0.6)+
  th+
  theme(
    axis.text.x = element_text(angle = 0),
    panel.grid.minor = element_blank())

dos <- ggplot(schools)+
  geom_boxplot(aes(gm,I_porc_mat-I_porc_len,fill=tipo_escuela),alpha=0.6)+
  th+
  theme(
    axis.text.x = element_text(angle = 0),
    panel.grid.minor = element_blank())

grid.arrange(unos, dos, ncol=1)
```

