

Introdução

Esse estudo é realizado tendo como base de dados o cadastro de CNPJ's (Cadastro Nacional de Pessoas Jurídicas), mas referente aqueles que

```
# carregando os arquivos
empresas = read.csv('bq-empresas.csv')
#filtrando para retirar entidades públicas e candidatos
nat_jud = qsacnpj::tab_natureza_juridica
nat_jud$cod_subclass_natureza_juridica = as.numeric(nat_jud$cod_subclass_natureza_juridica)
empresas = empresas|>
  dplyr::left_join(nat_jud, by = c("natureza_juridica" = "cod_subclass_natureza_juridica"))|>
  dplyr::filter(nm_natureza_juridica %in% c("Entidades Empresariais"))|>
  dplyr::distinct()

estabelecimentos = read.csv('bq-estabelecimentos.csv')
estabelecimentos = estabelecimentos|>
  dplyr::semi_join(empresas, by = "cnpj_basico")

socios = read.csv('bq-socios.csv')
# qsacnpj::tab_cnpj_entres_publicos_br
responsvel = qsacnpj::tab_qualificacao_responsavel_socio
sit_cad = qsacnpj::tab_situacao_cadastral
nat_jud = qsacnpj::tab_natureza_juridica
empresas_pi_atual = estabelecimentos|>
  dplyr::filter(situacao_cadastral == 2,
    identificador_matriz_filial == 1)
```

Início e fim das atividades das empresas no Piauí por Ano

```
#início das atividade
inicio_atividade = estabelecimentos|>
  dplyr::filter(identificador_matriz_filial == 1)|>
  dplyr::mutate(data_inicio_atividade = ymd(data_inicio_atividade))|>
  dplyr::mutate(mes_atividade = floor_date(data_inicio_atividade, "quarter"),
    ano_atividade = floor_date(data_inicio_atividade, "year"))|>
  dplyr::filter(mes_atividade >= "1954-01-01",
    mes_atividade <= "2022-10-01")|>
  count(ano_atividade)|>
  dplyr::mutate(mes_atividade = ano_atividade)|>
  filter(year(mes_atividade) >= 1963,
    year(mes_atividade) <= 2022)

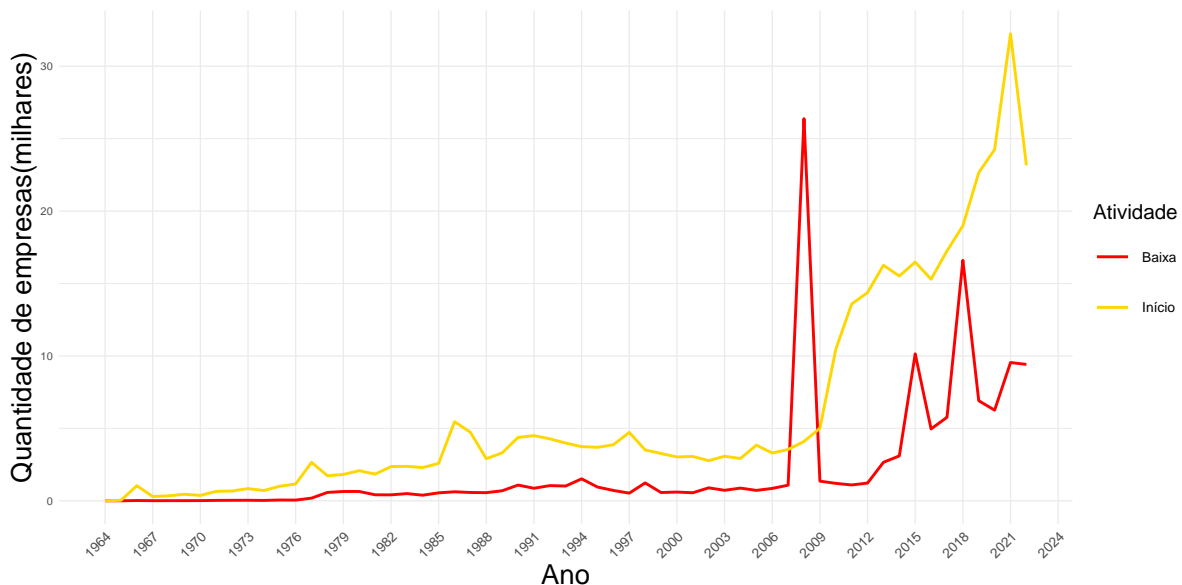
# fim das atividade
fim_atividade = estabelecimentos|>
  dplyr::filter(situacao_cadastral == 8, identificador_matriz_filial == 1)|>
  dplyr::mutate(data_situacao_cadastral = ymd(data_situacao_cadastral))|>
  dplyr::mutate(mes_atividade = floor_date(data_situacao_cadastral, "quarter"),
    ano_atividade = floor_date(data_situacao_cadastral, "year"))|>
  dplyr::filter(mes_atividade >= "1954-01-01",
    mes_atividade <= "2022-10-01")|>
```

```

count(ano_atividade)|>
dplyr::mutate(mes_atividade = ano_atividade)|>
filter(year(mes_atividade) >= 1963,
       year(mes_atividade) <= 2022)

#gráfico comparativo
inicio_atividade|>
left_join(fim_atividade, by = "mes_atividade")|>
dplyr::mutate(n.y = replace_na(n.y, 0))|>
dplyr::select(mes_atividade, n.x, n.y)|>
reshape2::melt(id = c("mes_atividade"))|>
dplyr::mutate(variable = str_replace(variable, "n.x", "Início"),
              variable = str_replace(variable, "n.y", "Baixa"))|>
dplyr::rename("Atividade" = "variable")|>
ggplot(aes(x = mes_atividade, y = value/1000, color = Atividade))+
  geom_line(size=1)+
  theme_minimal(10)+
  scale_color_manual(values = c("red", "gold"))+
  scale_x_date(breaks = scales::date_breaks("3 year"),
              labels = scales::date_format("%Y"))+
  labs(x = "Ano",
       y = "Quantidade de empresas(milhares)") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1, size=10),
        axis.title.x = element_text(size=20),
        axis.title.y = element_text(size=20),
        legend.key.size = unit(1.3, 'cm'),
        legend.key.width = unit(1, "cm"),
        legend.title = element_text(size=15),
        legend.text = element_text(size=10),
        panel.grid.minor.x = element_blank())

```



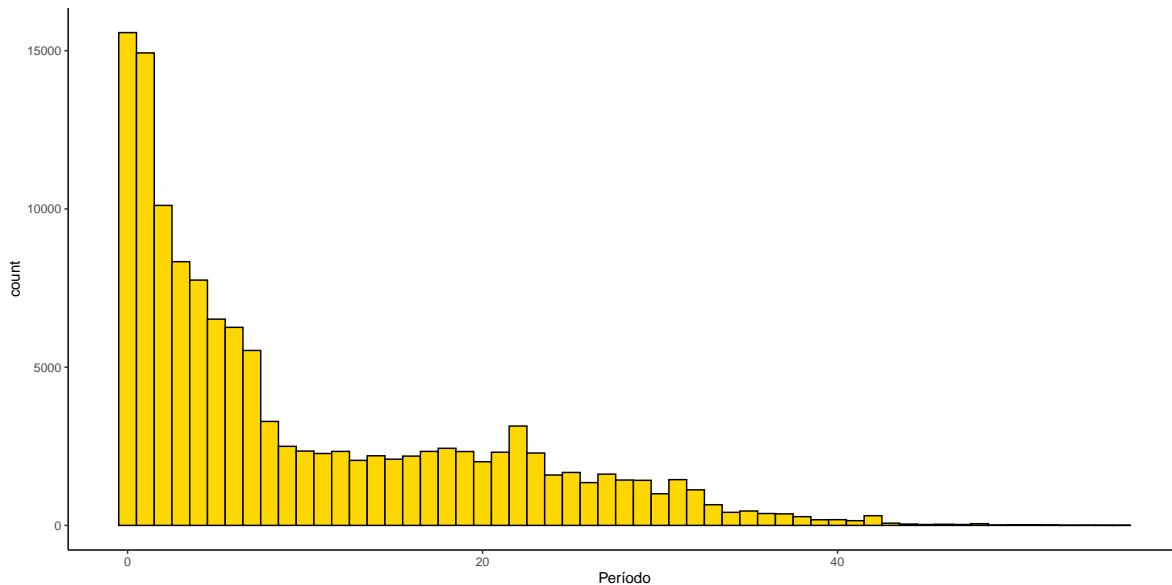
- motivo do número de cancelamentos em 2008: <https://crc-se.jusbrasil.com.br/noticias/3148127/pessoas->

juridicas-inaptas-terao-seu-cnpj-cancelado. No dia 31/12/2008, ocorreu a baixa de 25595 cnpts no Piauí.

Tempo de funcionamento, em anos, de empresas que tiveram baixa no cnpj

```
periodo =
estabelecimentos|>
  dplyr::mutate(data_situacao_cadastral = ymd(data_situacao_cadastral),
                data_inicio_atividade = ymd(data_inicio_atividade))|>
  dplyr::filter(situacao_cadastral == 8,
                identificador_matriz_filial == 1)|>
  dplyr::mutate(Período = lubridate::interval(data_inicio_atividade, data_situacao_
dplyr::select(Período)|>
  dplyr::mutate(Período = round(as.numeric(Período), 2))

ggplot(periodo, aes(x = Período))+
  geom_histogram(binwidth=1, colour = "black", fill = "gold")+
  theme_classic()
```



```
summary(periodo$Período)
```

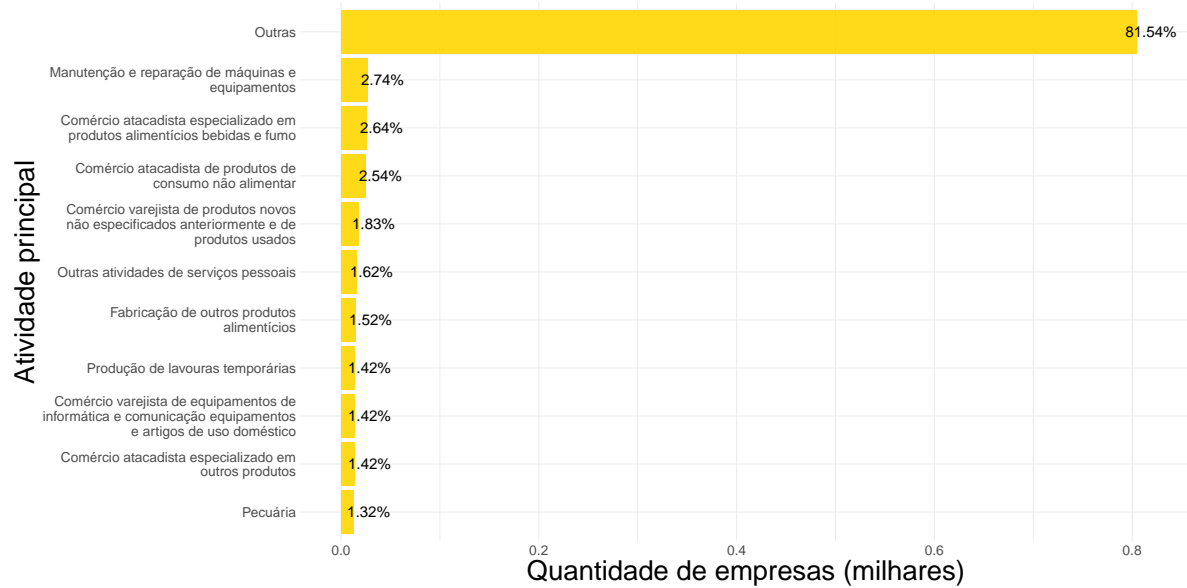
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.00	1.67	5.71	9.86	16.82	56.10

Tipo de empresa em atividade

Existem alguns ítem no comercio de reparação de veículos que não podem classificados como de veículo, utiliza outra classificação, a CNAE

```
cnae = qsacnpj::tab_cnae
cnae$cod_cnae = as.integer(cnae$cod_cnae)
nm_division = empresas_pi_atual|>
  dplyr::rename("cod_cnae" = "cnae_fiscal_principal")|>
  left_join(cnae, "cod_cnae")|>
  count(cod_cnae, nm_secao, nm_divisao, nm_grupo, nm_classe, nm_cnae)|>
  replace_na(list(nm_secao = "Vazio", nm_divisao = "Vazio",
                 nm_grupo = "Vazio", nm_classe = "Vazio",
                 nm_cnae = "Vazio"))

nm_division|>
  dplyr::mutate(tipo = fct_lump(nm_grupo, 10, other_level = "Outras")) %>%
  dplyr::count(tipo) %>%
  dplyr::mutate(tipo = fct_reorder(str_wrap(tipo, 38), n))|>
  dplyr::mutate(prop = percent(n/sum(n)))|>
  ggplot(aes(x = tipo, y = n/1e3))+
  geom_col(fill = 'gold', alpha = .9, width = 0.9) +
  coord_flip() +
  geom_text(aes(label = prop), nudge_y = 0.015) +
  theme_minimal(7.25) +
  labs(x = "Atividade principal",
       y = "Quantidade de empresas (milhares)") +
  theme(axis.title.x = element_text(size=20),
        axis.title.y = element_text(size=20),
        axis.text.x = element_text(size = 10),
        axis.text.y = element_text(size = 10))
```



Treemap para atividades de industria de transformação

```
#install.packages("treemapify", dependencies = T)
library(treemapify)

nm_division|>
  dplyr::filter(nm_secao == 'INDÚSTRIAS DE TRANSFORMAÇÃO')|>
  dplyr::mutate(tipo = fct_lump(nm_divisao, 11, other_level = "Outras")) %>%
  count(tipo)|>
  dplyr::mutate(tipo = fct_reorder(str_wrap(tipo, 20), n))|>
  ggplot(aes(area = n, fill = tipo, label = tipo))+
  geom_treemap()+
  geom_treemap_text(colour = "black",
                    place = "topleft",
                    size = 12) +
  theme(legend.position = "none")
```



Treemap para atividades de COMÉRCIO REPARAÇÃO DE VEÍCULOS AUTOMOTORES E MOTOCICLETAS

```
nm_division|>
  dplyr::filter(nm_secao ==
                 'COMÉRCIO REPARAÇÃO DE VEÍCULOS AUTOMOTORES E MOTOCICLETAS')|>
  dplyr::mutate(tipo = fct_lump(nm_grupo ,8, other_level = "Outras")) |>
  count(tipo)|>
  dplyr::mutate(tipo = fct_reorder(str_wrap(tipo, 20), n))|>
  ggplot(aes(area = n, fill = tipo, label = tipo))+
  geom_treemap()+
  geom_treemap_text(
    place = "topleft",
    size = 15)+
  scale_fill_brewer(palette = "YlOrBr")+
  theme(legend.position = "none")
```

Comércio atacadista especializado em produtos alimentícios bebidas e fumo	Representantes comerciais e agentes do comércio exceto de veículos automotores e motocicletas	Comércio varejista de produtos alimentícios bebidas e fumo
Outras	Comércio varejista de equipamentos de informática e comunicação equipamentos e artigos de uso doméstico	Comércio varejista de material de construção
	Comércio varejista de produtos novos não especificados anteriormente e de produtos usados	Comércio atacadista especializado em outros produtos
	Comércio atacadista de produtos de consumo não alimentar	

Empresas por Cidade

```
library(geobr)

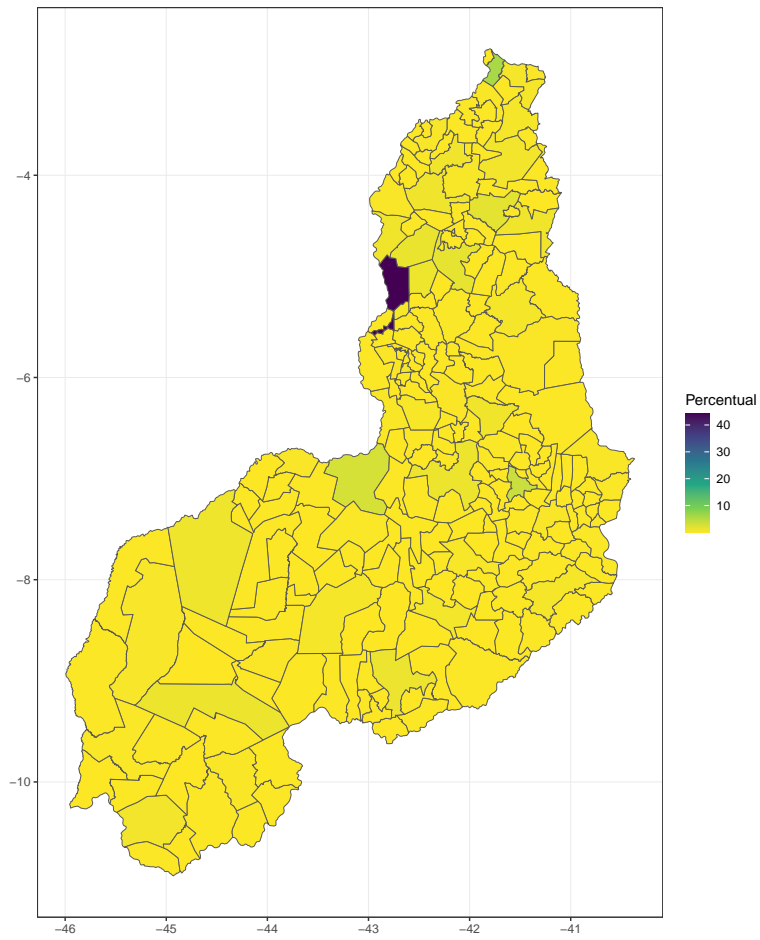
all_mun_pi <- read_municipality(code_muni=22, year=2018)

dates_pi = empresas_pi_atual|>
  group_by(id_municipio)|>
  summarise("Quantidade" = n())|>
  dplyr::mutate("Percentual" = percent(Quantidade*100/sum(Quantidade)))|>
  #dplyr::rename("code_muni" = "id_municipio")|>
  dplyr::left_join(all_mun_pi, by=c("id_municipio"= "code_muni"))

min_empresas = min(dates_pi$Quantidade)
max_empresas = max(dates_pi$Quantidade)
min_percentual = min(dates_pi$Percentual)
max_percentual = max(dates_pi$Percentual)

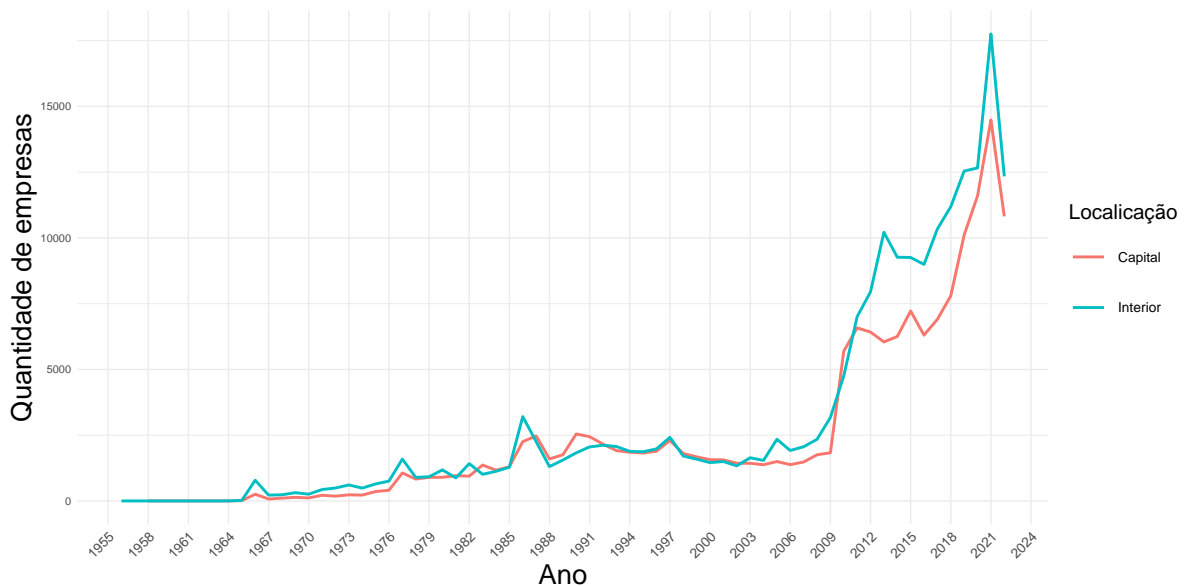
ggplot() +
  geom_sf(data = dates_pi, aes(fill=Percentual, geometry = geom))+
  scale_fill_viridis_c(direction = -1, limits = c(min_percentual,max_percentual))+
  labs(title= "Empresas por cidade - 2022",
       subtitle = "Percentual") +
  theme_bw()
```

Empresas por cidade – 2022
Percentual



Abertura de empresas entre interior e Capital

```
estabelecimentos|>
  #left_join(empresas, "cnpj_basico")|>
  dplyr::filter(identificador_matriz_filial == 1)|>
  dplyr::mutate(data_inicio_atividade = ymd(data_inicio_atividade))|>
  dplyr::mutate(mes_atividade = floor_date(data_inicio_atividade, "quarter"),
               ano_atividade = floor_date(data_inicio_atividade, "year"),
               Localização = case_when(id_municipio == 2211001 ~ "Capital",
                                       id_municipio != 2211001 ~ "Interior"))|>
  dplyr::filter(mes_atividade >= "1954-01-01",
               mes_atividade <= "2022-10-01")|>
  count(ano_atividade, Localização)|>
  na.omit()|>
  ggplot(aes(x = ano_atividade, y = n, color = Localização))+
  geom_line(size=1)+
  theme_minimal(10)+
  scale_x_date(breaks = scales::date_breaks("3 year"),
              labels = scales::date_format("%Y"))+
  labs(x = "Ano",
       y = "Quantidade de empresas") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1, size=10),
        axis.title.x = element_text(size=20),
        axis.title.y = element_text(size=20),
        legend.key.size = unit(1.3, 'cm'),
        legend.key.width = unit(1, "cm"),
        legend.title = element_text(size=15),
        legend.text = element_text(size=10),
        panel.grid.minor.x = element_blank())
```



Bairros com mais empresas em Teresina

- Obs.: Existem muitos nomes de bairros repetidos mas com gráfia diferente, o que gera uma lista muito grande. Usamos funções que retiram os acentos e espaços a esquerda ou direita para diminuir as diferenças.

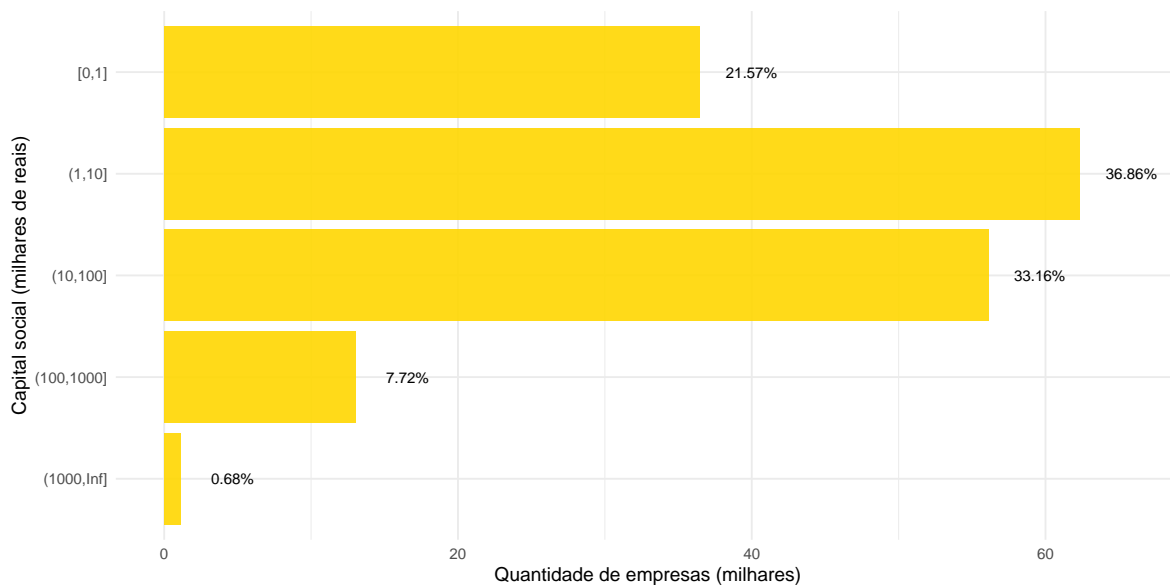
```
empresas_pi_atual|>
  filter(id_municipio == 2211001)|>
  dplyr::mutate(bairro = stringi::stri_trans_general(str = bairro, id = "Latin-ASCII"))|>
  dplyr::mutate(bairro = str_trim(bairro, "right"),# remove espaços a direita
               bairro = str_trim(bairro, "left"))|> # remove espaços a esquerda
  dplyr::mutate(Região = fct_lump(bairro ,20, other_level = "Outras"))|>
  count(Região)|>
  arrange(desc(n))|>
  knitr::kable()
```

Região	n
Outras	39117
CENTRO	6849
ITARARE	3207
ANGELIM	2061
JOQUEI	2020
MOCAMBINHO	1928
FATIMA	1880
VALE QUEM TEM	1748
LOURIVAL PARENTE	1717
SAO CRISTOVAO	1570
VERMELHA	1366
PARQUE PIAUI	1304
ININGA	1265
SANTO ANTONIO	1252
PROMORAR	1198
GURUPI	1121
PARQUE IDEAL	1062
NOVO HORIZONTE	1012
ESPLANADA	974
SANTA MARIA	955
RENASCENCA	951

Capital Social

```
empresas_pi_atual|>
  left_join(empresas, "cnpj_basico")|>#View()
  dplyr::mutate(capital_social_cat = cut(capital_social/1e3,
                                         breaks = c(0, 1e3, 1e4, 1e5, 1e6, Inf)/1e3,
                                         include.lowest = TRUE, dig.lab = 10))|>

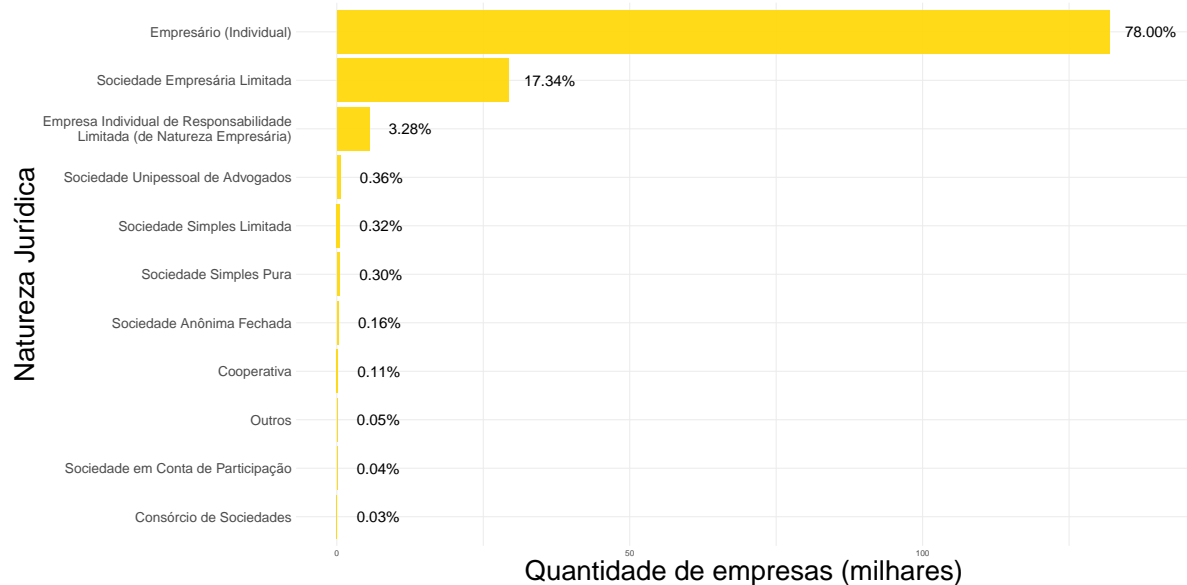
  count(tipo = capital_social_cat) |>
  dplyr::mutate(prop = percent(n/sum(n)),
                tipo = fct_rev(tipo))|>
  ggplot(aes(x = tipo, y = n / 1e3)) +
  geom_col(fill = "gold", alpha = .9) +
  coord_flip() +
  geom_text(aes(label = prop), nudge_y = 3.5) +
  theme_minimal(14) +
  labs(x = "Capital social (milhares de reais)",
       y = "Quantidade de empresas (milhares)")
```



Natureza jurídica

```
nat_jud$cod_subclass_natureza_juridica = as.integer(nat_jud$cod_subclass_natureza_juridica)

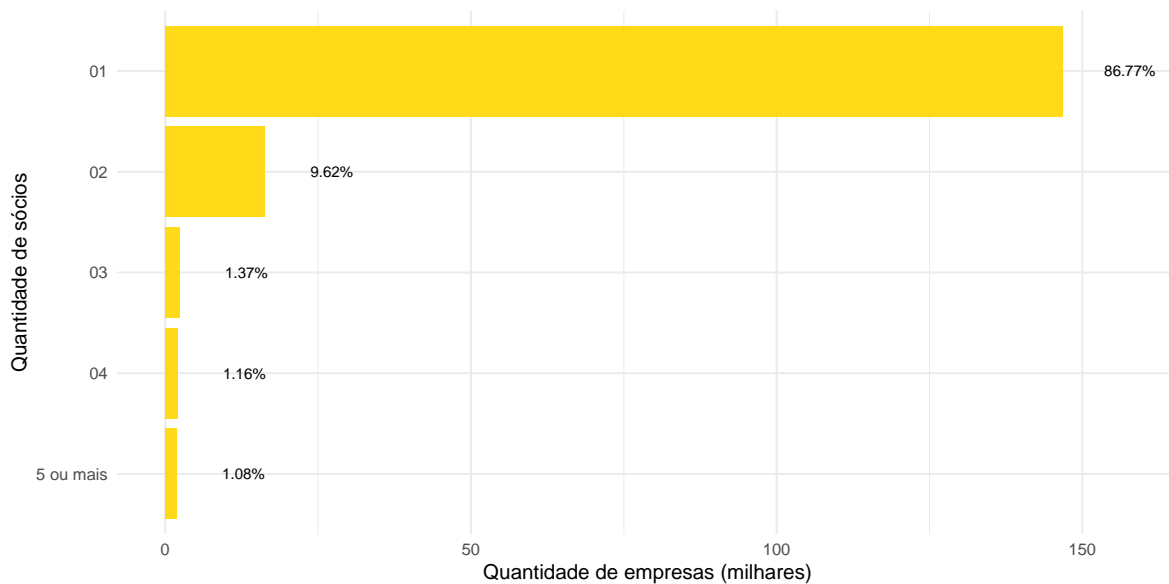
empresas_pi_atual|>
  left_join(empresas, "cnpj_basico")|>
  replace_na(
    list(nm_natureza_juridica = "Vazio",
         nm_subclass_natureza_juridica = "Vazio"))|>
  mutate(tipo = if_else(
    nm_natureza_juridica == "Administração Pública",
    "Administração Pública", nm_subclass_natureza_juridica
  ),
  tipo = fct_lump(tipo, 10, other_level = "Outros"))|>
  count(tipo)|>
  dplyr::mutate(prop = percent(n/sum(n)),
               tipo = fct_reorder(str_wrap(tipo, 38), n))|>
  ggplot(aes(x = tipo, y = n/ 1e3)) +
  geom_col(fill = "gold", alpha = .9, width = 0.9) +
  coord_flip() +
  geom_text(aes(label = prop), nudge_y = 7) +
  theme_minimal(7.25) +
  labs(x = "Natureza Jurídica",
       y = "Quantidade de empresas (milhares)") +
  theme(axis.title.x = element_text(size=20),
        axis.title.y = element_text(size=20),
        axis.text.y = element_text(size = 10),
        )
```



Quantidade de Sócios

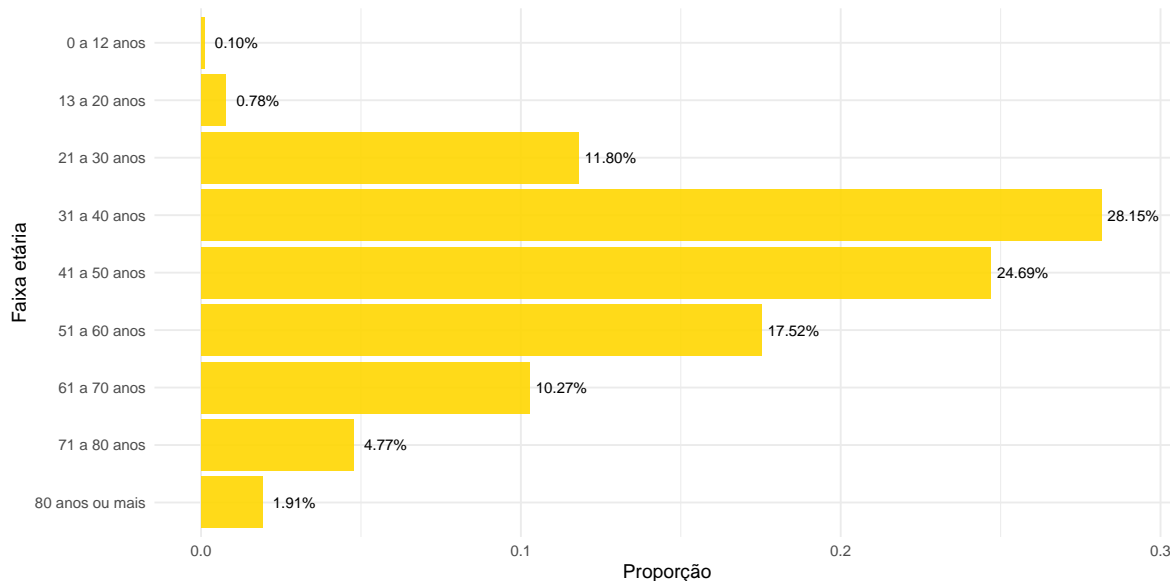
```
join_socios_empresas = empresas_pi_atual|>
  left_join(socios, "cnpj_basico")

join_socios_empresas|>
  count(cnpj_basico)|>
  dplyr::mutate(n = fct_lump(str_pad(n, 2, "left", 0), 4,
                             other_level = "5 ou mais"))|>
  dplyr::mutate(n= fct_lump(n, 10, other_level = "11 ou mais"))|>
  count(n) |>
  mutate(prop = percent(nn/sum(nn)), n = fct_rev(n)) |>
  ggplot(aes(x = n, y = nn / 1e3)) +
  geom_col(fill = "gold", alpha = .9) +
  coord_flip() +
  geom_text(aes(label = prop), nudge_y = 11) +
  theme_minimal(15) +
  labs(x = "Quantidade de sócios",
       y = "Quantidade de empresas (milhares)")
```



Idade dos Sócios

```
join_socios_empresas|>
  dplyr::filter(!is.na(faixa_etaria))|>
  dplyr::filter(faixa_etaria > 0)|>
  dplyr::mutate(Idade = case_when(
    faixa_etaria == 1 ~ "0 a 12 anos",
    faixa_etaria == 2 ~ "13 a 20 anos",
    faixa_etaria == 3 ~ "21 a 30 anos",
    faixa_etaria == 4 ~ "31 a 40 anos",
    faixa_etaria == 5 ~ "41 a 50 anos",
    faixa_etaria == 6 ~ "51 a 60 anos",
    faixa_etaria == 7 ~ "61 a 70 anos",
    faixa_etaria == 8 ~ "71 a 80 anos",
    faixa_etaria == 9 ~ "80 anos ou mais" ))|>
  count(Idade)|>
  dplyr::mutate(prop = percent(n/sum(n)),
    Idade = fct_rev(Idade))|>
  ggplot(aes(x = Idade, y = prop)) +
  geom_col(fill = "gold", alpha = .9) +
  coord_flip() +
  geom_text(aes(label = prop), nudge_y = 0.01) +
  theme_minimal(14) +
  labs(x = "Faixa etária",
    y = "Proporção")
```



Sexo dos proprietários:

obs.: na tabela não tem o sexo do proprietários, então fizemos uma pesquisa de acordo com o nome na tabela sócios usando o pacote `get_gender` e função `genderBR`.

```
# Pegando primeiro nome
teste = sapply(strsplit(as.character(join_socios_empresas$nome), ' '), "[", 1)

# contando os valores ausentes
table(is.na(teste))
```

```
FALSE  TRUE
82636 132360
```

```
# deletando valores NA
teste = na.omit(teste)

# presumindo o sexo
sexo = genderBR::get_gender(teste)

# contando valores ausentes sexo
table(is.na(sexo))
```

```
FALSE  TRUE
76649  5987
```

```
# deletando valores NA
sexo = na.omit(sexo)

# determinando o sexo
sexo|>table()|>prop.table()|>percent()
```

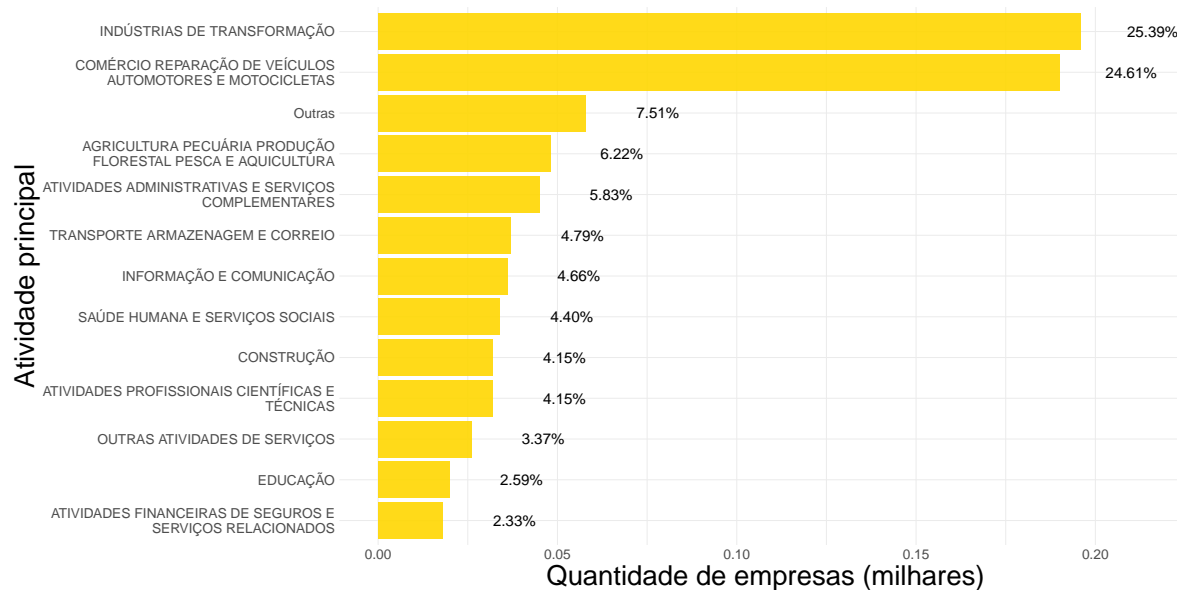
```
sexo
Female  Male
  37%   63%
```

Atividades reunião 05/12/22

Atividades fechadas

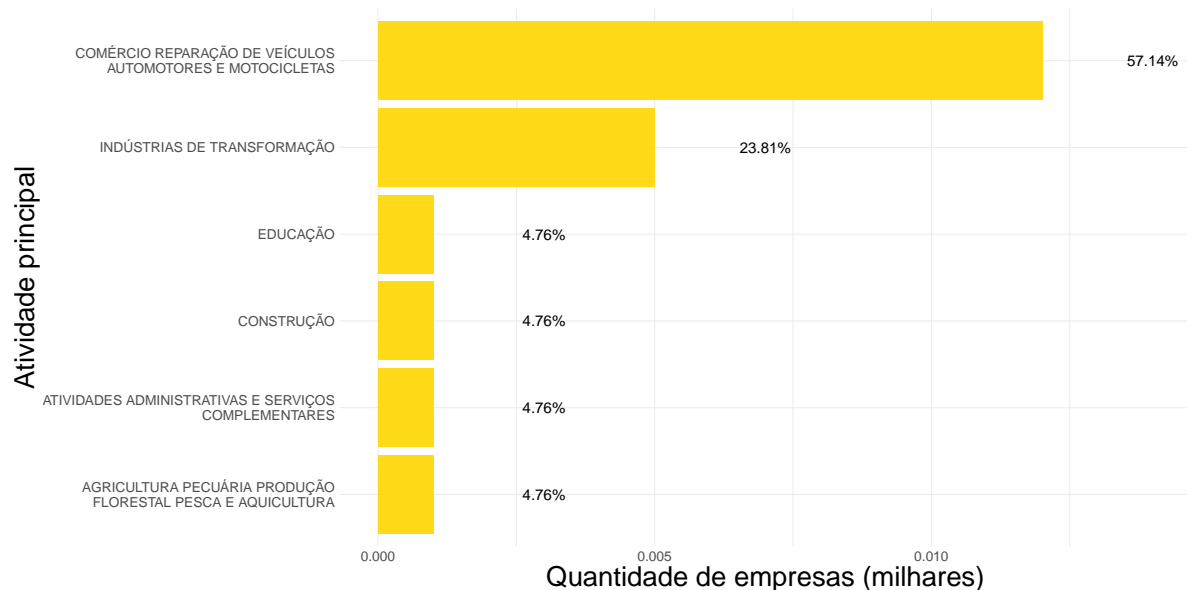
```
empresas_5 = estabelecimentos|>
  dplyr::filter(identificador_matriz_filial == 1,
                data_situacao_cadastral >= '2017-10-01')

empresas_5|>
  dplyr::filter(situacao_cadastral == 8)|>
  dplyr::rename("cod_cnae" = "cnae_fiscal_principal")|>
  left_join(cnae, "cod_cnae")|>
  count(nm_secao, nm_divisao, nm_grupo, nm_classe, nm_cnae)|>
  replace_na(list(nm_secao = "Vazio", nm_divisao = "Vazio",
                 nm_grupo = "Vazio", nm_classe = "Vazio",
                 nm_cnae = "Vazio"))|>
  dplyr::mutate(tipo = fct_lump(nm_secao, 12, other_level = "Outras")) %>%
  dplyr::count(tipo) %>%
  dplyr::mutate(tipo = fct_reorder(str_wrap(tipo, 38), n))|>
  dplyr::mutate(prop = percent(n/sum(n)))|>
  ggplot(aes(x = tipo, y = n/1e3)) +
  geom_col(fill = 'gold', alpha = .9, width = 0.9) +
  coord_flip() +
  geom_text(aes(label = prop), nudge_y = 0.020) +
  theme_minimal(7.25) +
  labs(x = "Atividade principal",
       y = "Quantidade de empresas (milhares)") +
  theme(axis.title.x = element_text(size=20), axis.title.y = element_text(size=20),
        axis.text.x = element_text(size = 10), axis.text.y = element_text(size = 10))
```



Atividades abertas

```
empresas_5|>
  dplyr::filter(situacao_cadastral == 1)|>
  dplyr::rename("cod_cnae" = "cnae_fiscal_principal")|>
  left_join(cnae, "cod_cnae")|>
  count(nm_secao, nm_divisao, nm_grupo, nm_classe, nm_cnae)|>
  replace_na(list(nm_secao = "Vazio", nm_divisao = "Vazio",
    nm_grupo = "Vazio", nm_classe = "Vazio",
    nm_cnae = "Vazio"))|>
  dplyr::mutate(tipo = fct_lump(nm_secao, 12, other_level = "Outras")) %>%
  dplyr::count(tipo) %>%
  dplyr::mutate(tipo = fct_reorder(str_wrap(tipo, 38), n))|>
  dplyr::mutate(prop = percent(n/sum(n)))|>
  ggplot(aes(x = tipo, y = n/1e3)) +
  geom_col(fill = 'gold', alpha = .9, width = 0.9) +
  coord_flip() +
  geom_text(aes(label = prop), nudge_y = 0.0020) +
  theme_minimal(7.25) +
  labs(x = "Atividade principal",
    y = "Quantidade de empresas (milhares)") +
  theme(axis.title.x = element_text(size=20), axis.title.y = element_text(size=20),
    axis.text.x = element_text(size = 10), axis.text.y = element_text(size = 10))
```



Relação Natureza jurídica e Tipo de atividade

- não é importante realizar essa relação entre natureza e tipo de atividade, como estamos trabalhando com o criação de empresas, as naturezas jurídicas serão *entidades empresariais*

```
empresas_5|>
  left_join(empresas, "cnpj_basico")|>
  #left_join(nat_jud, by=c("natureza_juridica"= "cod_subclass_natureza_juridica"))|>
  replace_na(list(nm_natureza_juridica = "Vazio",
                  nm_subclass_natureza_juridica = "Vazio")) |>
  mutate(tipo = if_else(nm_natureza_juridica == "Administração Pública",
                        "Administração Pública", nm_subclass_natureza_juridica))|>
  count(tipo, nm_natureza_juridica)|>
  dplyr::mutate(prop = percent(n/sum(n)))|>
  arrange(desc(n))|>
  knitr::kable()
```

tipo	nm_natureza_juridica	n	prop
Empresário (Individual)	Entidades Empresariais	180340	83.72%
Sociedade Empresária Limitada	Entidades Empresariais	28693	13.32%
Empresa Individual de Responsabilidade Limitada (de Natureza Empresária)	Entidades Empresariais	4677	2.17%
Sociedade Unipessoal de Advogados	Entidades Empresariais	550	0.26%
Cooperativa	Entidades Empresariais	251	0.12%
Sociedade Simples Limitada	Entidades Empresariais	250	0.12%
Sociedade Simples Pura	Entidades Empresariais	250	0.12%
Sociedade Anônima Fechada	Entidades Empresariais	190	0.09%
Sociedade Empresária em Nome Coletivo	Entidades Empresariais	69	0.03%
Sociedade em Conta de Participação	Entidades Empresariais	40	0.02%
Empresa Individual de Responsabilidade Limitada (de Natureza Simples)	Entidades Empresariais	38	0.02%
Consórcio de Sociedades	Entidades Empresariais	30	0.01%
Sociedade Anônima Aberta	Entidades Empresariais	9	0.00%
Empresa Pública	Entidades Empresariais	3	0.00%
Cooperativas de Consumo	Entidades Empresariais	2	0.00%
Sociedade de Economia Mista	Entidades Empresariais	2	0.00%
Consórcio de Empregadores	Entidades Empresariais	1	0.00%

tipo	nm_natureza_juridica	n	prop
Sociedade Empresária em Comandita por Ações	Entidades Empresariais	1	0.00%
Sociedade Empresária em Comandita Simples	Entidades Empresariais	1	0.00%

Natureza, atividade por bairro de Teresina Ver também o capital social..

Quantidade de empresas por Região da Cidade

```
bairros = "bairros.xlsx"|>
  readxl::read_xlsx()|>
  dplyr::mutate(Bairro = stringi::stri_trans_general(str = Bairro, id = "Latin-ASC
    Bairro = str_trim(Bairro, "right"),# remove espaços a direita
    Bairro = str_trim(Bairro, "left"), # remove espaço a esquerda
    Bairro = toupper(Bairro)) # bota todas as letras em caixas alta

empresa_bairro_the = empresas_pi_atual|>
  filter(id_municipio == 2211001)|>
  dplyr::mutate(bairro = stringi::stri_trans_general(str = bairro, id
    bairro = str_trim(bairro, "right"),# remove espaços a
    bairro = str_trim(bairro, "left"))|> # remove espaços
  dplyr::left_join(bairros, by = c("bairro" = "Bairro"))|>
  dplyr::distinct()|>
  drop_na(Região)|> #6643 NA's
  count(Região)|>
  dplyr::mutate(prop = percent(n/sum(n)))

empresa_bairro_the|>knitr::kable()
```

Região	n	prop
Central	10635	15.65%
Zona Leste	16380	24.11%
Zona Norte	10664	15.70%
Zona Sudeste	10259	15.10%
Zona Sul	20002	29.44%

Média por Bairro

```
empresas_pi_atual|>
  left_join(empresas, "cnpj_basico")|>
  filter(id_municipio == 2211001)|>
  dplyr::mutate(bairro = stringi::stri_trans_general(str = bairro, id = "Latin-ASCII")
               bairro = str_trim(bairro, "right"), # remove espaços a direita
               bairro = str_trim(bairro, "left"))|> # remove espaços a esquerda
  dplyr::left_join(bairros, by = c("bairro" = "Bairro"))|>
  group_by(Região)|>
  summarise(`Média Capital` = mean(capital_social), Máximo = max(capital_social))|>
  knitr::kable()
```

Região	Média Capital	Máximo
Central	633815.15	1252570644
Zona Leste	613341.50	383436551
Zona Norte	35008.58	40000000
Zona Sudeste	69776.02	86579150
Zona Sul	139787.97	400000000
NA	122635.14	41000000

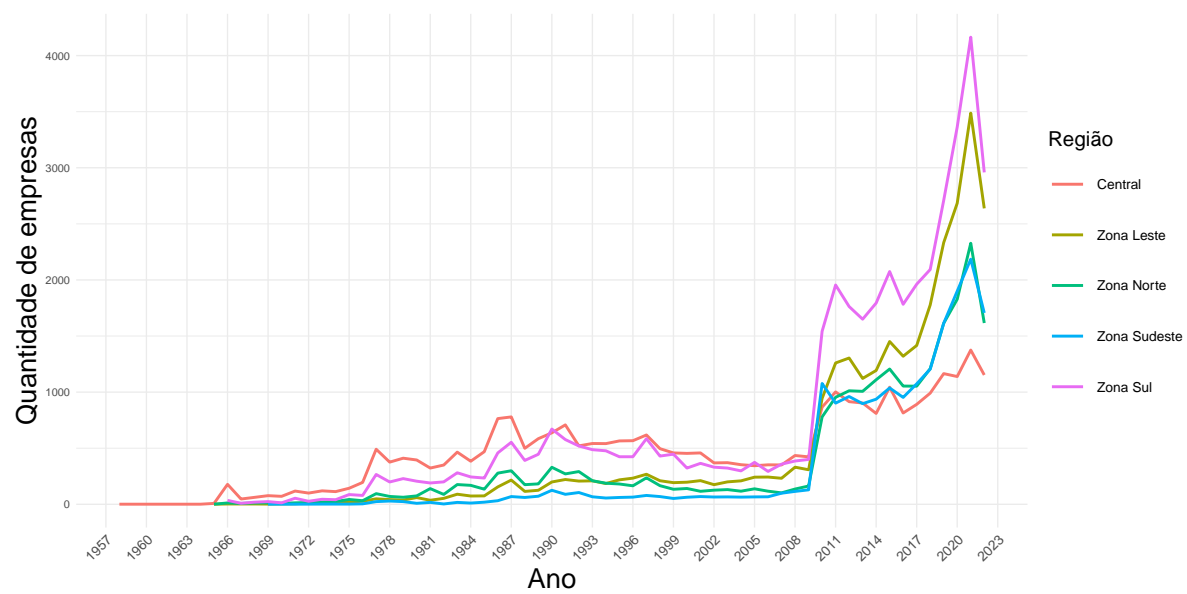
```
empresas_pi_atual|>
  left_join(empresas, "cnpj_basico")|>
  filter(id_municipio == 2211001)|>
  dplyr::mutate(bairro = stringi::stri_trans_general(str = bairro, id = "Latin-ASCII")
               bairro = str_trim(bairro, "right"), # remove espaços a direita
               bairro = str_trim(bairro, "left"))|> # remove espaços a esquerda
  dplyr::left_join(bairros, by = c("bairro" = "Bairro"))|>
  dplyr::arrange(desc(capital_social))|>
  select(razao_social, capital_social, bairro, Região)|>
  head(30)|>knitr::kable()
```

razao_social	capital_social	bairro	Região
CLAUDINO S A LOJAS DE DEPARTAMENTOS	1252570644	CENTRO	Central
HUMANA ASSISTENCIA MEDICA LTDA	982818816	CENTRO	Central
CLAUDINO PATRIMONIAL S.A.	836258352	CENTRO	Central
RDS PARTICACOES SOCIETARIAS LTDA	430214000	CENTRO	Central
HOSPITAL MED IMAGEM S.A	427280471	CENTRO	Central
FERRONORTE INDUSTRIAL LTDA	400000000	PEDRA MIUDA	Zona Sul
ENEL GREEN POWER VENTOS DE SAO ROQUE 01 S.A.	383436551	SAO CRISTOVAO	Zona Leste
ENEL GREEN POWER VENTOS DE SAO ROQUE 02 S.A.	369758651	SAO CRISTOVAO	Zona Leste
ENEL GREEN POWER VENTOS DE SAO ROQUE 16 S.A.	353284551	SAO CRISTOVAO	Zona Leste
ENEL GREEN POWER VENTOS DE SAO ROQUE 04 S.A.	339980531	SAO CRISTOVAO	Zona Leste
ENEL GREEN POWER VENTOS DE SAO ROQUE 08 S.A.	337473759	SAO CRISTOVAO	Zona Leste

razao_social	capital_social	bairro	Região
ENEL GREEN POWER VENTOS DE SAO ROQUE 18 S.A.	332473759	SAO CRISTOVAO	Zona Leste
ENEL GREEN POWER VENTOS DE SAO ROQUE 11 S.A.	318740451	SAO CRISTOVAO	Zona Leste
ENEL GREEN POWER VENTOS DE SANTA ANGELA 2 S.A.	299922006	SAO CRISTOVAO	Zona Leste
EQUATORIAL PIAUI DISTRIBUIDORA DE ENERGIA S.A	287479127	CENTRO	Central
CREDI-SHOP S/A - INSTITUICAO DE PAGAMENTO	259404525	ILHOTAS	Central
ENEL GREEN POWER VENTOS DE SAO ROQUE 17 S.A.	258952101	SAO CRISTOVAO	Zona Leste
SC2 SHOPPING RIO POTY LTDA	217213591	PORENQUANTO	Central
RDAMASIO IMOVEIS LTDA	214000000	CENTRO	Central
ENEL GREEN POWER VENTOS DE SANTA ANGELA 14 S.A.	198554956	SAO CRISTOVAO	Zona Leste
HOUSTON DO NORDESTE SA	186574200	DISTRITO INDUSTRIAL	Zona Sul
ENEL GREEN POWER SAO GONCALO 6 S.A.	183602691	SAO CRISTOVAO	Zona Leste
ENEL GREEN POWER VENTOS DE SANTA ANGELA 1 S.A.	182273006	SAO CRISTOVAO	Zona Leste
EMPREENDIMENTOS FARMACEUTICOS GLOBO LTDA	163020328	SAO CRISTOVAO	Zona Leste
BIKE DO NORDESTE S/A	154709167	DISTRITO INDUSTRIAL	Zona Sul
ENEL GREEN POWER VENTOS DE SANTA ANGELA 17 S.A.	152022288	SAO CRISTOVAO	Zona Leste
ELETRO DO NORDESTE S/A	150000000	PROMORAR	Zona Sul
ENEL GREEN POWER SAO GONCALO 14 S.A.	147279288	SAO CRISTOVAO	Zona Leste
SOCIMOL INDUSTRIA DE COLCHOES E MOVEIS S.A	144275902	TABULETA	Zona Sul
ENEL GREEN POWER VENTOS DE SANTA ANGELA 11 S.A.	132786606	SAO CRISTOVAO	Zona Leste

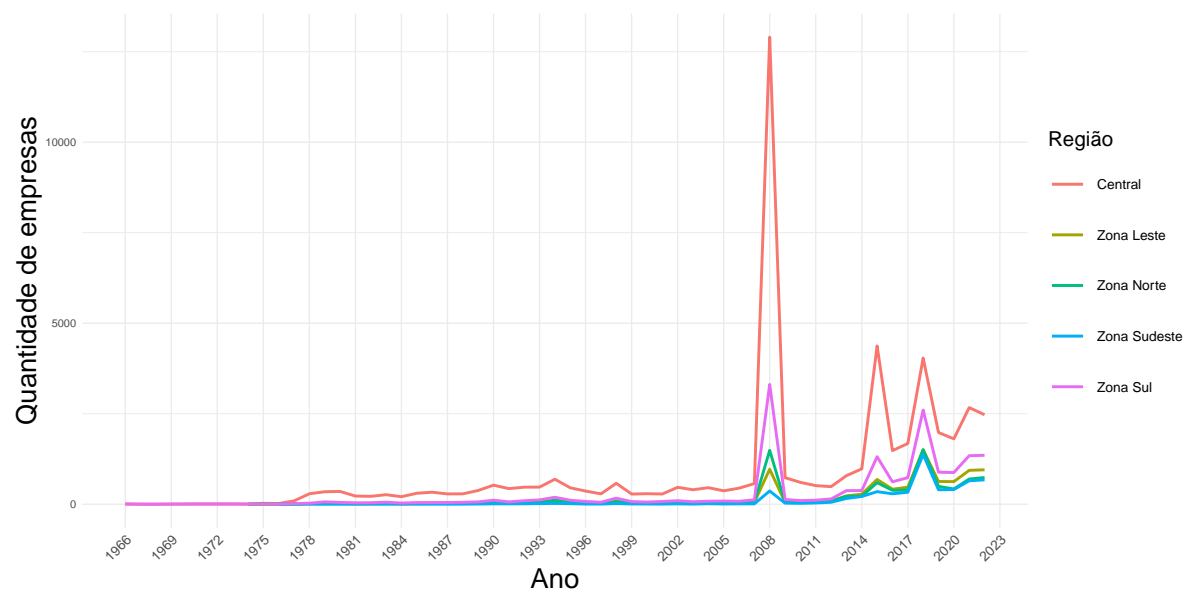
Evolução das empresas por região da cidade

```
estabelecimentos|>
  #left_join(empresas, "cnpj_basico")|>
  filter(id_municipio == 2211001)|>
  dplyr::mutate(bairro = stringi::stri_trans_general(str = bairro, id = "Latin-ASCII"),
               bairro = str_trim(bairro, "right"), # remove espaços a direita
               bairro = str_trim(bairro, "left"))|> # remove espaços a esquerda
  dplyr::left_join(bairros, by = c("bairro" = "Bairro"))|>
  dplyr::filter(identificador_matriz_filial == 1)|>
  dplyr::mutate(data_inicio_atividade = ymd(data_inicio_atividade))|>
  dplyr::mutate(mes_atividade = floor_date(data_inicio_atividade, "quarter"),
               ano_atividade = floor_date(data_inicio_atividade, "year"))|>
  dplyr::filter(mes_atividade >= "1954-01-01",
               mes_atividade <= "2022-10-01")|>
  count(ano_atividade, Região)|>
  na.omit())|>
  ggplot(aes(x = ano_atividade, y = n, color = Região))+
  geom_line(size=1)+
  theme_minimal(10)+
  scale_x_date(breaks = scales::date_breaks("3 year"),
              labels = scales::date_format("%Y"))+
  labs(x = "Ano",
       y = "Quantidade de empresas") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1, size=10),
        axis.title.x = element_text(size=20),
        axis.title.y = element_text(size=20),
        legend.key.size = unit(1.3, 'cm'),
        legend.key.width = unit(1, "cm"),
        legend.title = element_text(size=15),
        legend.text = element_text(size=10),
        panel.grid.minor.x = element_blank())
```



Fechamento de empresas

```
estabelecimentos|>
  left_join(empresas, "cnpj_basico")|>
  #filter(id_municipio == 2211001)|>
  dplyr::mutate(bairro = stringi::stri_trans_general(str = bairro, id = "Latin-ASCII"),
               bairro = str_trim(bairro, "right"), # remove espaços a direita
               bairro = str_trim(bairro, "left"))|> # remove espaços a esquerda
  dplyr::left_join(bairros, by = c("bairro" = "Bairro"))|>
  dplyr::filter(situacao_cadastral == 8, identificador_matriz_filial == 1)|>
  dplyr::mutate(data_situacao_cadastral = ymd(data_situacao_cadastral))|>
  dplyr::mutate(mes_atividade = floor_date(data_situacao_cadastral, "quarter"),
               ano_atividade = floor_date(data_situacao_cadastral, "year"))|>
  dplyr::filter(mes_atividade >= "1954-01-01",
               mes_atividade <= "2022-10-01")|>
  count(ano_atividade, Região)|>
  na.omit()|>
  ggplot(aes(x = ano_atividade, y = n, color = Região))+
  geom_line(size=1)+
  theme_minimal(10)+
  scale_x_date(breaks = scales::date_breaks("3 year"),
               labels = scales::date_format("%Y"))+
  labs(x = "Ano",
       y = "Quantidade de empresas") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1, size=10),
        axis.title.x = element_text(size=20),
        axis.title.y = element_text(size=20),
        legend.key.size = unit(1.3, 'cm'),
        legend.key.width = unit(1, "cm"),
        legend.title = element_text(size=15),
        legend.text = element_text(size=10),
        panel.grid.minor.x = element_blank())
```



- Pico do ano de 2008 foi explicado anteriormente.

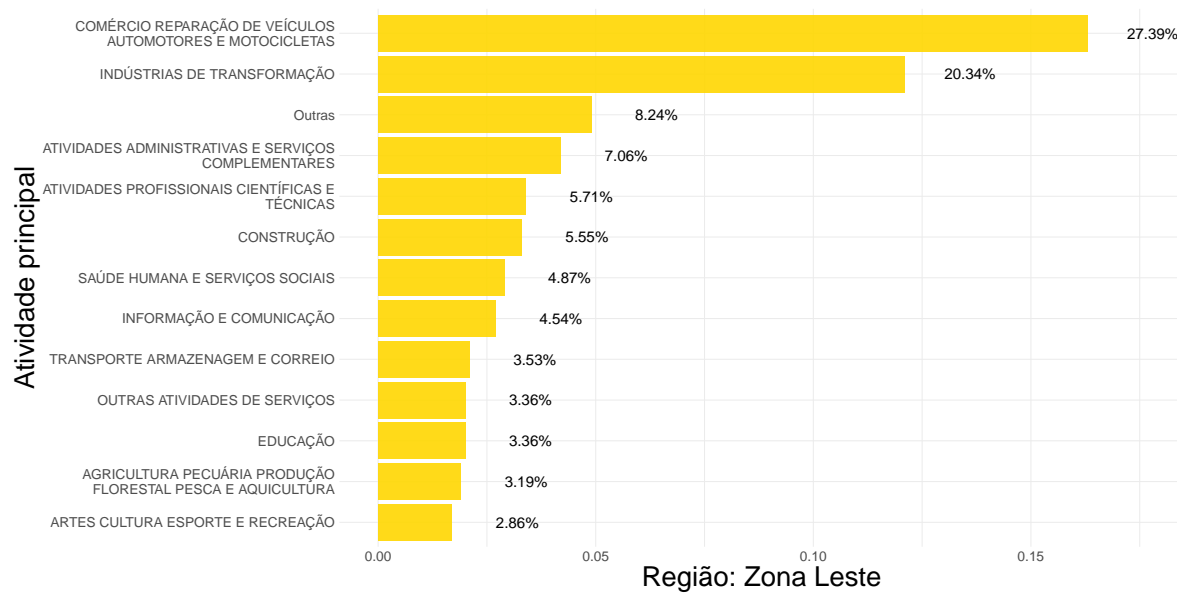
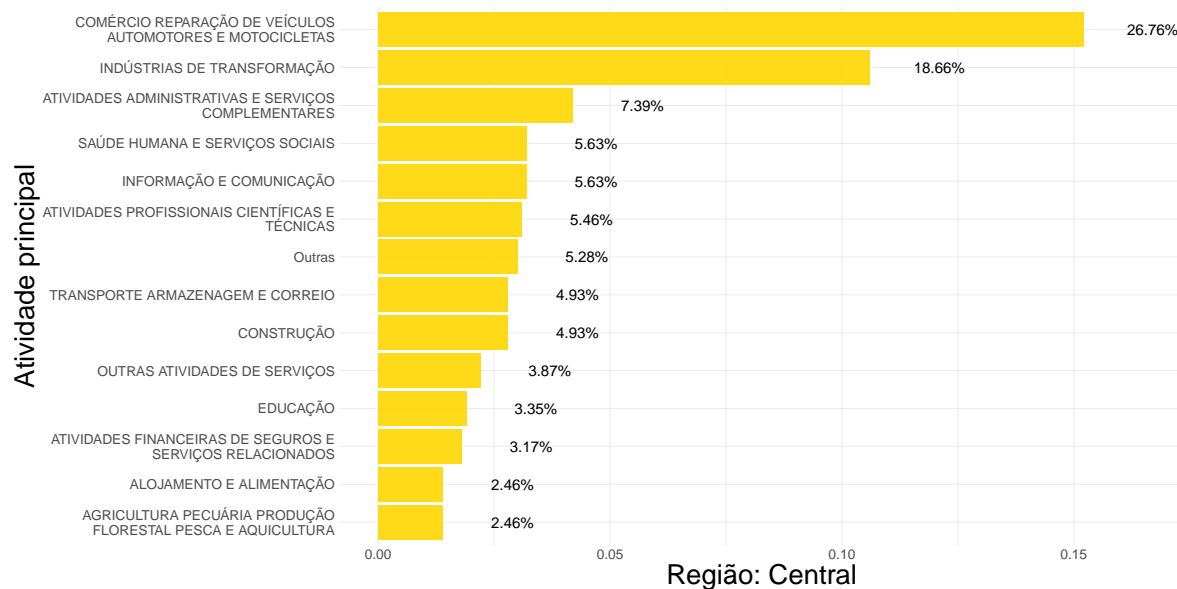
Atividade por Região

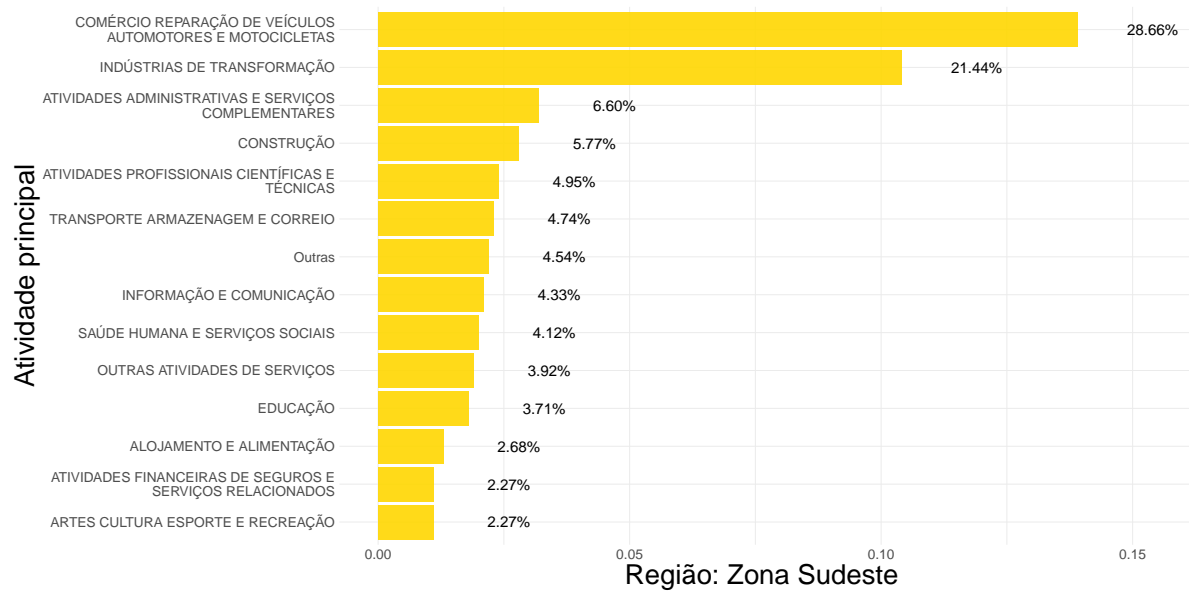
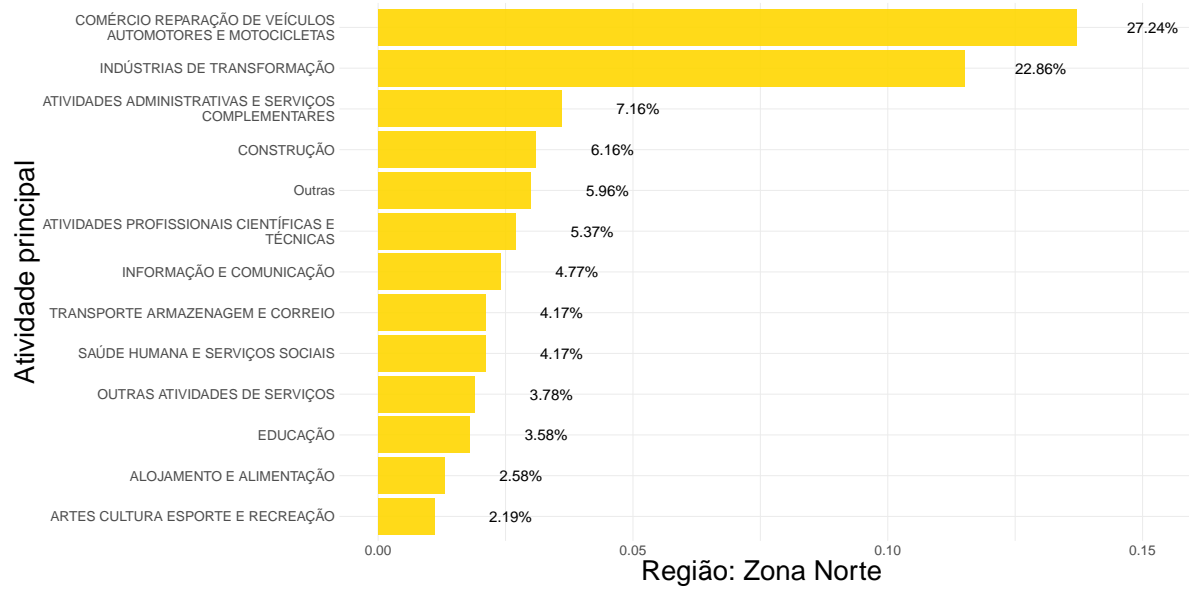
```
write.csv(nm_division, "cnae_the.csv")

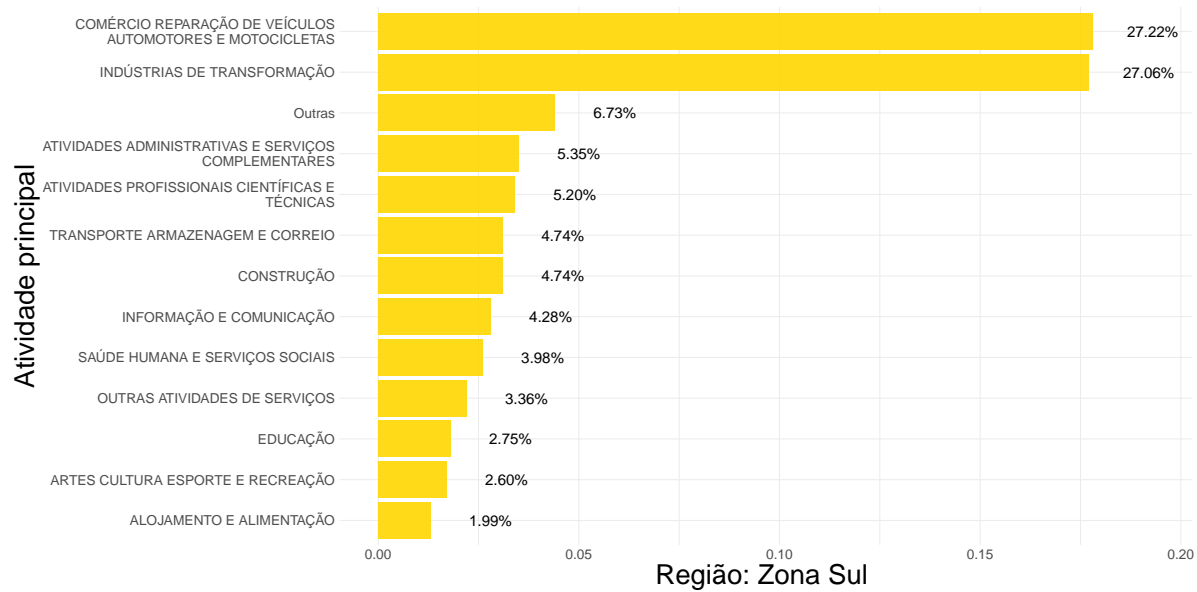
empresa_regiao = function(zona){
  empresas_pi_atual|>
  left_join(empresas, "cnpj_basico")|>
  filter(id_municipio == 2211001)|>
  dplyr::mutate(bairro = stringi::stri_trans_general(str = bairro, id = "Latin-ASCII"),#
               bairro = str_trim(bairro, "right"),# remove espaços a direita
               bairro = str_trim(bairro, "left"))|> # remove espaços a esquerda
  dplyr::left_join(bairros, by = c("bairro" = "Bairro"))|>
  dplyr::filter(Região == zona)|>
  dplyr::rename("cod_cnae" = "cnae_fiscal_principal")|>
  left_join(cnae, "cod_cnae")|>
  count(nm_secao, nm_divisao, nm_grupo, nm_classe, nm_cnae)|>
  replace_na(list(nm_secao = "Vazio", nm_divisao = "Vazio",
                 nm_grupo = "Vazio", nm_classe = "Vazio",
                 nm_cnae = "Vazio"))|>
  dplyr::mutate(tipo = fct_lump(nm_secao,12, other_level = "Outras")) %>%
  dplyr::count(tipo) %>%
  dplyr::mutate(tipo = fct_reorder(str_wrap(tipo, 38), n))|>
  dplyr::mutate(prop = percent(n/sum(n)))|>
  ggplot(aes(x = tipo, y = n/1e3))+
  geom_col(fill = 'gold', alpha = .9, width = 0.9) +
  coord_flip() +
  geom_text(aes(label = prop), nudge_y = 0.015) +
  theme_minimal(7.25) +
  labs(x = "Atividade principal",
       y = sprintf("Região: %s ", zona))+
  theme(axis.title.x = element_text(size = 20),
        axis.title.y = element_text(size = 20),
        axis.text.x = element_text(size = 10),
        axis.text.y = element_text(size = 10))
}

zonas = c("Central", "Zona Leste", "Zona Norte", "Zona Sudeste", "Zona Sul")

for (i in zonas){
  print(empresa_regiao(i))
}
```







Referências

- <https://blog.curso-r.com/posts/2019-09-20-qsacnpj/>
- <https://github.com/georgevbsantiago/qsacnpj>
- https://basedosdados.org/dataset/br-me-cnpj?bdm_table=empresas
- <https://www.gov.br/receitafederal/pt-br/assuntos/orientacao-tributaria/cadastros/consultas/dados-publicos-cnpj>