

# Questão 3

Wellington Charles Lacerda Nobrega

17/04/2021

## Limpeza dos Dados

```
# limpar o cachê
rm(list = ls())

# configuração de exibição de decimais
options(digits = "2", scipen = 99)

# Opção para lidar com a Unidade Primária de Amostragem
options(survey.lonely.psu = "adjust")

# definição do diretório do arquivo a ser lido
path_pnad <- paste0(getwd(), "/pnad.csv")

# Carregando os Pacotes Necessários
ip <- installed.packages()
for (i in c("tidyverse", "readxl", "ggplot2", "stargazer", "xtable", "gridExtra", "knitr")){
  if ( !(i %in% ip) ) { install.packages(i) }
  if (i=="Quandl") { library(i, character.only = T) ; Quandl.auth("WP2rt8HsRo3kjWsRkLY5") }
  else library(i, character.only = T)
}
rm(list=c("ip", "i"))

# lendo os dados e colocando o path por segurança
pnad <- read.csv(path_pnad, sep = ",")

# Limpando a base de dados
dados <- pnad %>%
  # retirando as pessoas que nasceram no dia 99
  filter(V2008 != 99) %>%
  # identificando o domicílio
  unite(col = "domicilio", c(UPA, V1008, V1014), sep = "-", remove = TRUE) %>%
  # criando uma coluna de data de nascimento
  unite(col = "data_nascimento", c(V2008, V20081, V20082), sep = "-", remove = TRUE) %>%
  # criando uma variável indivíduo
  unite(col = "indivíduo", c(domicilio, data_nascimento, V2007, Ano), sep = "-", remove = FALSE) %>%
  # criando uma variável para denotar os gêmeos
  mutate(gemeos = ifelse(duplicated(indivíduo) | duplicated(indivíduo, fromLast = TRUE), 1, 0)) %>%
  # retirando os gêmeos
  filter(gemeos != 1) %>% select(-gemeos)
```

```
# removendo a base não filtrada da memória do R
rm(pnad)
```

## 1. Análise das Pessoas Ocupadas

```
# selecionando os grupos que vou utilizar de acordo com o solicitado 2020
ocupados_2019 <- dados %>% filter(Ano == 2019 & VD4002 == 1)
ocupados_2020 <- dados %>% filter(Ano == 2020 & VD4002 == 1)

# número de pessoas ocupadas em 2019 e 2020
o19 <- ocupados_2019 %>% group_by(Ano, VD4002) %>% summarise(n())
o20 <- ocupados_2020 %>% group_by(Ano, VD4002) %>% summarise(n())

# número de pessoas na força de trabalho em 2019 e 2020
f19 <- dados %>% filter(Ano==2019, VD4001==1) %>% group_by(Ano) %>% summarise(n())
f20 <- dados %>% filter(Ano==2020, VD4001==1) %>% group_by(Ano) %>% summarise(n())

tab1 <- cbind(
  rbind( "Taxa de Desocupação (%)" = (1 - (o19$n()/f19$n()))*100, "Nº ocupados" = o19$n(), "Força de Trabalho" = f19$n()),
  rbind( "Taxa de Desocupação (%)" = (1 - (o20$n()/f20$n()))*100, "Nº ocupados" = o20$n(), "Força de Trabalho" = f20$n()),
  rbind( "VAR% 19/20" = 100*((1 - (o20$n()/f20$n()))*100 - (1 - (o19$n()/f19$n()))*100) / ((1 - (o19$n()/f19$n()))*100),
        "VAR% pop.des. 19/20" = ((o20$n() - o19$n())/(o19$n()))*100, "VAR% for.tra.19/20" = ((f20$n() - f19$n())/f19$n())*100)

colnames(tab1) <- c("2019", "2020", "VAR%")

# removendo objetos que não serão mais usados
rm(o19,o20,f19,f20)

# visualização dos resultados
knitr::kable(tab1, caption = "Taxa de Desocupação, Número de Ocupados e Força de Trabalho")
```

Table 1: Taxa de Desocupação, Número de Ocupados e Força de Trabalho

	2019	2020	VAR%
Taxa de Desocupação (%)	10	13	22
Nº ocupados	218261	122187	-44
Força de Trabalho	243514	139876	-43

```
horas <- dados %>% filter(VD4001==1) %>% group_by(Ano) %>% summarise("media_horas_habitual"=mean(VD4031),
"media_horas_efetiva"=mean(VD4035,na.rm=TRUE))

horas <- data.frame(horas)
knitr::kable(horas, align = T, caption = "Média de Horas trabalhadas (Habitual e Efetiva)")
```

Table 2: Média de Horas trabalhadas (Habitual e Efetiva)

Ano	media_horas_habitual	media_horas_efetiva
2019	39	38
2020	39	37

```

# idade dos ocupados
idade      <- dados %>% filter(VD4002 == 1) %>% select(Ano,V2009,V2007)
idade2019  <- ocupados_2019 %>% select(V2009,V2007)
idade2020  <- ocupados_2019 %>% select(V2009,V2007)

# sexo dos ocupados
gen2019 <- ocupados_2019 %>% group_by(V2007) %>% summarise(n())
gen2020 <- ocupados_2020 %>% group_by(V2007) %>% summarise(n())

# plotando a pirâmide populacional por gênero dos ocupados 2019/2020

labs <- c(paste(seq(0, 85, by = 5), seq(0 + 5 - 1, 90 - 1, by = 5),
              sep = "-"), paste(90, "+", sep = ""))

idade      <- dados %>% filter(VD4002 == 1) %>% mutate(V2007 = if_else(V2007 == 1,"Homem", "Mulher")) %>%
idade$GrupoEtario <- cut(idade$V2009,breaks = c(seq(0, 90, by = 5), Inf), labels = labs, right = FALSE)
idade <- idade %>% group_by(Ano, V2007, GrupoEtario) %>% summarise("total"=n()) %>%
  mutate(sinal = ifelse(V2007 == "Homem",1,-1)) %>% mutate(total2 = total*sinal)

plot1 <- grid.arrange(
  ggplot(idade, aes(x = GrupoEtario, y = total2, fill = V2007)) +
    ggtitle("Demografia dos Ocupados 2019") +
    labs(y = "Milhares") +
    geom_bar(data = subset(idade, Ano == 2019 & V2007 == "Homem"),stat= "identity") +
    geom_bar(data = subset(idade, Ano == 2019 & V2007 == "Mulher"), stat= "identity") +
    scale_y_continuous(breaks = seq(-15000, 15000, 1000),
      labels = paste0(as.character(c(15:0, 1:15)), "")) +
    scale_fill_brewer(palette = "Set1") + coord_flip() +
    theme(panel.background = element_rect(fill = "white", colour = "grey25"),
      legend.background = element_rect(fill=NA),
      legend.position=c(0.85,0.9),
      legend.key = element_rect(fill = "white", colour = "white"))
  ,
  ggplot(idade, aes(x = GrupoEtario, y = total2, fill = V2007)) +
    ggtitle("Demografia dos Ocupados 2020") +
    labs(y = "Milhares") +
    geom_bar(data = subset(idade, Ano == 2020 & V2007 == "Homem"),stat= "identity") +
    geom_bar(data = subset(idade, Ano == 2020 & V2007 == "Mulher"), stat= "identity") +
    scale_y_continuous(breaks = seq(-15000, 15000, 1000),
      labels = paste0(as.character(c(15:0, 1:15)), "")) +
    scale_fill_brewer(palette = "Set1") +
    coord_flip() +
    theme(panel.background = element_rect(fill = "white", colour = "grey25"),
      legend.background = element_rect(fill=NA),
      legend.position=c(0.85,0.9),
      legend.key = element_rect(fill = "white", colour = "white"))
  , ncol = 2)

# analisando a relação entre faixa etária e população ocupada

idade      <- dados %>% filter(VD4002 == 1) %>% mutate(V2007 = if_else(V2007 == 1,"Homem", "Mulher")) %>%

```

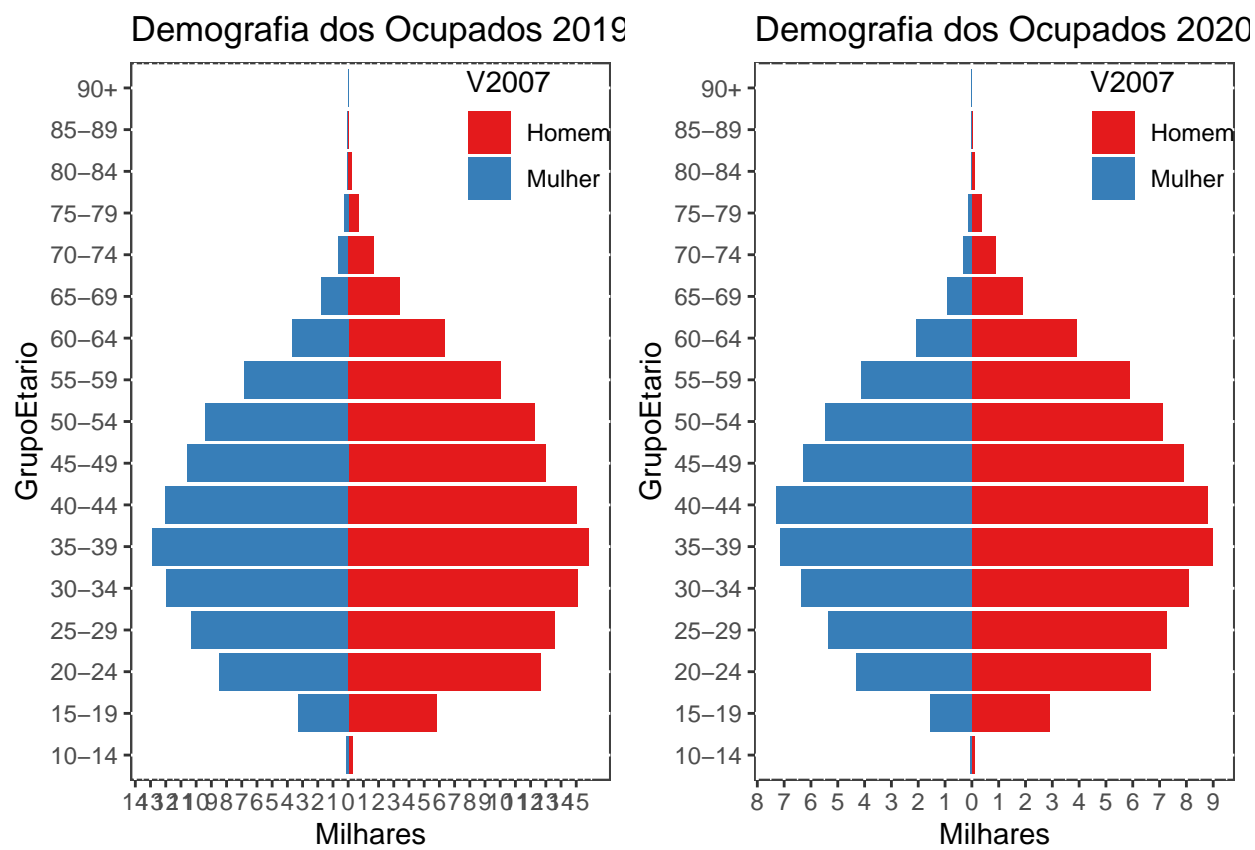


Figure 1: Pirâmide demográfica dos indivíduos ocupados

```

idade$GrupoEtario <- cut(idade$V2009,breaks = c(seq(0, 90, by = 5), Inf), labels = labs, right = FALSE)
idade <- idade %>% group_by(Ano, V2007, GrupoEtario) %>% summarise("total"=n())

perh19 <- idade %>% filter(Ano == 2019, V2007 == "Homem") %>%
  mutate( pp = (total/sum(total)*100))
perh20 <- idade %>% filter(Ano == 2020, V2007 == "Homem") %>%
  mutate( pp = (total/sum(total)*100))
perm19 <- idade %>% filter(Ano == 2019, V2007 == "Mulher") %>%
  mutate( pp = (total/sum(total)*100))
perm20 <- idade %>% filter(Ano == 2020, V2007 == "Mulher") %>%
  mutate( pp = (total/sum(total)*100))

labs1 <- c(paste(seq(10, 85, by = 5), seq(10 + 5 - 1, 90 - 1, by = 5),
  sep = "-"), paste(90, "+", sep = ""))

part_ocupada19 <- cbind("Faixa Etária"=labs1,"% homem 19" = round(perh19$pp,2), "% mulher 19" = round(perm19$pp,2))
part_ocupada20 <- cbind("Faixa Etária"=labs1,"% homem 20" = round(perh20$pp,2), "% mulher 20" = round(perm20$pp,2))

tab2 <- cbind("Faixa Etária"=labs1,"% homem 19" = round(perh19$pp,2), "% mulher 19" = round(perm19$pp,2),
"% homem 20" = round(perh20$pp,2), "% mulher 20" = round(perm20$pp,2))

# visualização
knitr::kable(tab2, caption = "Participação (%) por Gênero e Idade na População Ocupada")

```

## Análise da Idade das Pessoas Ocupadas

Table 3: Participação (%) por Gênero e Idade na População Ocupada

Faixa Etária	% homem 19	% mulher 19	% homem 20	% mulher 20
10-14	0.22	0.15	0.16	0.11
15-19	4.62	3.58	4.08	3
20-24	10.04	9.16	9.41	8.39
25-29	10.77	11.17	10.23	10.42
30-34	11.97	12.95	11.4	12.37
35-39	12.55	13.93	12.67	13.88
40-44	11.95	13	12.43	14.18
45-49	10.32	11.48	11.14	12.26
50-54	9.75	10.17	10.04	10.63
55-59	7.97	7.41	8.31	8.05
60-64	5.06	4.01	5.49	4.01
65-69	2.67	1.91	2.67	1.78
70-74	1.31	0.69	1.25	0.62
75-79	0.55	0.29	0.52	0.22
80-84	0.19	0.07	0.15	0.07
85-89	0.05	0.03	0.05	0.01
90+	0.01	0	0.01	0.01

```
labs2 <- c("Sem_Instrução","Fund_Incom", "Fund_Comp","Médio_Incom","Médio_Comp","Sup_Incom","Sup_Comp")
```

```

# analisando a relação entre faixa etária e população ocupada
escolaridade_2019 <- dados %>% filter(VD4002 == 1 & Ano == 2019) %>% group_by(Ano, VD3004) %>% summaris
escolaridade_2020 <- dados %>% filter(VD4002 == 1 & Ano == 2020) %>% group_by(Ano, VD3004) %>% summaris

# criando uma tabela para apresentar os resultados
tab3 <- cbind(
  "Nível" = labs2,
  "Nº- 2019" = escolaridade_2019$n(),
  "% - 2019" = round((escolaridade_2019$n()/sum(escolaridade_2019$n()))*100,2),
  "Nº - 2020" = escolaridade_2020$n(),
  "% - 2020" = round((escolaridade_2020$n()/sum(escolaridade_2020$n()))*100,2)
)

# visualização
knitr::kable(tab3, aling = T, caption = "Escolaridade da População Ocupada (Total e %)")

```

## Análise da Escolaridade das Pessoas Ocupadas

Table 4: Escolaridade da População Ocupada (Total e %)

Nível	Nº- 2019	% - 2019	Nº - 2020	% - 2020
Sem_Instrução	5856	2.68	2837	2.32
Fund_Incom	57904	26.53	29194	23.89
Fund_Comp	17896	8.2	9609	7.86
Médio_Incom	14733	6.75	7744	6.34
Médio_Comp	69133	31.67	39024	31.94
Sup_Incom	12491	5.72	7386	6.04
Sup_Comp	40248	18.44	26393	21.6

```

escolaridade <- dados %>% filter(VD4002 == 1) %>% group_by(Ano, VD3004) %>% summarise("total"=n())
escolaridade$Nível <- cut(escolaridade$VD3004,breaks = c(seq(1,7 , by = 1), Inf), labels = labs2, right

plot2 <- ggplot(escolaridade, aes(x = Nível, y = total, fill = factor(Ano))) +
  ggtitle("Distribuição da Escolaridade dos Ocupados") +
  labs(y = "", x = "") +
  geom_bar(stat= "identity",position = position_dodge()) +
  #geom_bar(data = subset(escolaridade, Ano == 2019), stat= "identity",position = position_dodge()) +
  #geom_bar(data = subset(escolaridade, Ano == 2020), stat= "identity") +
  theme(panel.background = element_rect(fill = "white"),
        legend.background = element_rect(fill=NA),
        legend.position=c(0.85,0.9),
        legend.key = element_rect(fill = "white", colour = "white")) +
  geom_text(aes(label=total), position=position_dodge(width=0.9), vjust=-0.25) +
  scale_fill_brewer(palette = "Set1")

```

plot2

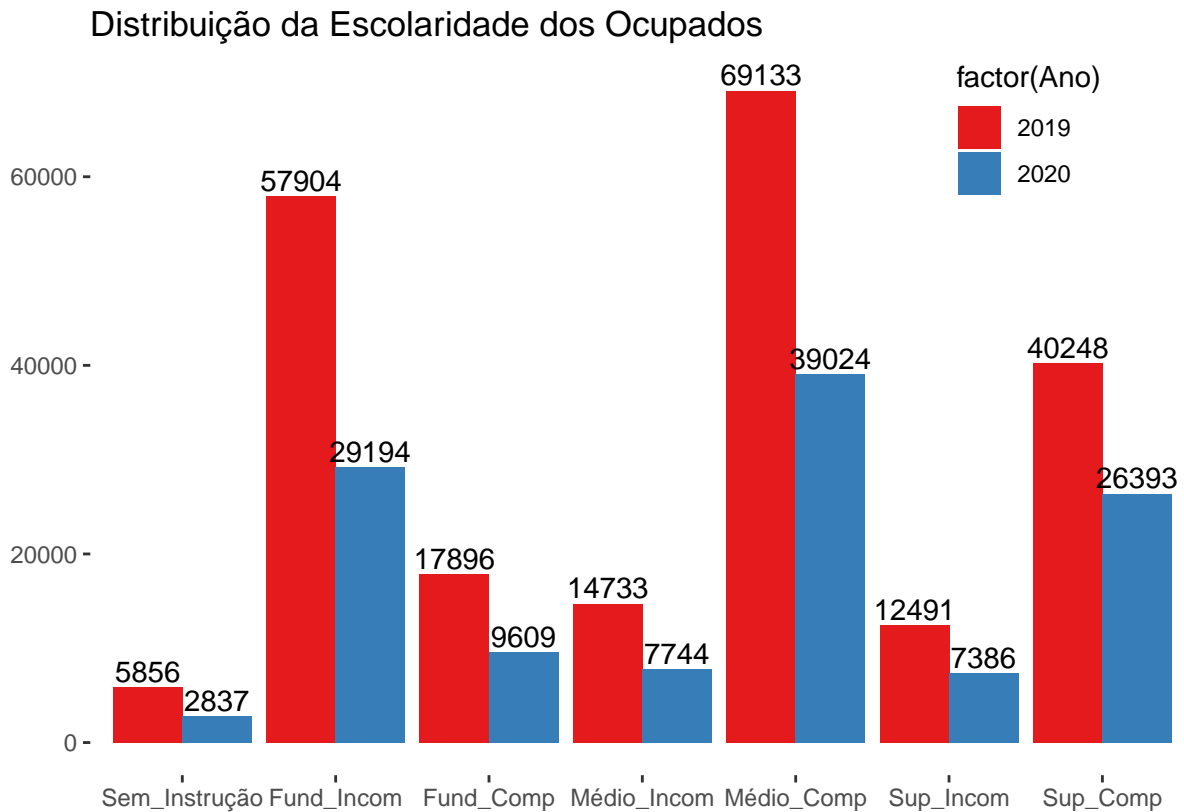


Figure 2: Distribuição da Escolaridade dos Ocupados

```
# analisando a relação entre faixa etária e população ocupada
raca_2019 <- dados %>% filter(VD4002 == 1 & Ano == 2019) %>% group_by(Ano, V2010) %>% summarise("total19" = sum(V2010))
raca_2020 <- dados %>% filter(VD4002 == 1 & Ano == 2020) %>% group_by(Ano, V2010) %>% summarise("total20" = sum(V2010))
labs3 <- c("branca", "preta", "amarela", "parda", "indigena", "ignorado")

# criando uma tabela para apresentar os resultados
tab4 <- cbind(
  "Nível" = labs3,
  "Nº - 2019" = raca_2019$total19,
  "% - 2019" = 100*((raca_2019$total19)/(sum(raca_2019$total19))),
  "Nº - 2020" = raca_2020$total20,
  "% - 2020" = ((raca_2020$total20)/(sum(raca_2020$total20)))*100
)

# visualização
knitr::kable(tab4, ailing = c, digits = 2, caption = "Distribuição da População Ocupada de Acordo com Cor ou Raça")
```

Distribuição da População Ocupada de Acordo com a Cor ou Raça

Table 5: Distribuição da População Ocupada de Acordo com Cor ou Raça (Total e %)

Nível	Nº- 2019	% - 2019	Nº - 2020	% - 2020
branca	89191	40.8643779694952	52063	42.6092792195569
preta	20923	9.58622933093865	10596	8.6719536448231
amarela	1199	0.54934230118986	623	0.509874209203925
parda	105933	48.5350108356509	58461	47.8455154803703
indigena	981	0.449461882791704	441	0.360922193031992
ignorado	34	0.0155776799336574	3	0.00245525301382307

```

raca <- dados %>% filter(VD4002 == 1) %>% group_by(Ano, V2010) %>% summarise("total"=n())
raca$factor <- cut(raca$V2010,breaks = c(seq(1,6, by = 1), Inf), labels = labs3, right = FALSE)

plot3 <- ggplot(raca, aes(x = factor, y = total, fill = factor(Ano))) +
  ggtitle("Distribuição por Raça da População Ocupada") +
  labs(y = "", x = "") +
  geom_bar(stat= "identity",position = position_dodge()) +
  #geom_bar(data = subset(escolaridade, Ano == 2019), stat= "identity",position = position_dodge()) +
  #geom_bar(data = subset(escolaridade, Ano == 2020), stat= "identity") +
  theme(panel.background = element_rect(fill = "white"),
        legend.background = element_rect(fill=NA),
        legend.position=c(0.85,0.9),
        legend.key = element_rect(fill = "white", colour = "white")) +
  geom_text(aes(label=total), position=position_dodge(width=0.9), vjust=-0.25) +
  scale_fill_brewer(palette = "Set1")

plot3

```

```

renda <- dados %>% filter(VD4001==1) %>% group_by(Ano) %>% summarise("renda_media_efetiva"=mean(VD4020,
"renda_media_habitual"=mean(VD4019,na.rm=TRUE))

renda$factor <- c("A","B")

plot4 <- grid.arrange(
  ggplot(renda) +
    geom_bar(aes(y = renda_media_efetiva, x = factor(Ano), fill=factor(Ano)),stat="identity",position = position_dodge()) +
    geom_text(aes(label=round(renda_media_efetiva,0), x = factor(Ano), y=renda_media_efetiva), position=position_dodge()) +
    theme(legend.position="none",
          panel.background = element_rect(fill = "white")) +
    labs(y = "R$", x = "Renda Média Efetiva")+ scale_fill_brewer(palette = "Set1")
  ,
  ggplot(renda) +
    geom_bar(aes(y = renda_media_habitual, x = factor(Ano), fill=factor(Ano)), stat="identity",position = position_dodge()) +
    geom_text(aes(label=round(renda_media_habitual,0), x = factor(Ano), y=renda_media_habitual), position=position_dodge()) +
    theme(legend.position="none",
          panel.background = element_rect(fill = "white")) +
    labs(y = "", x = "Renda Média Habitual") + scale_fill_brewer(palette = "Set1")
  , ncol = 2)

```



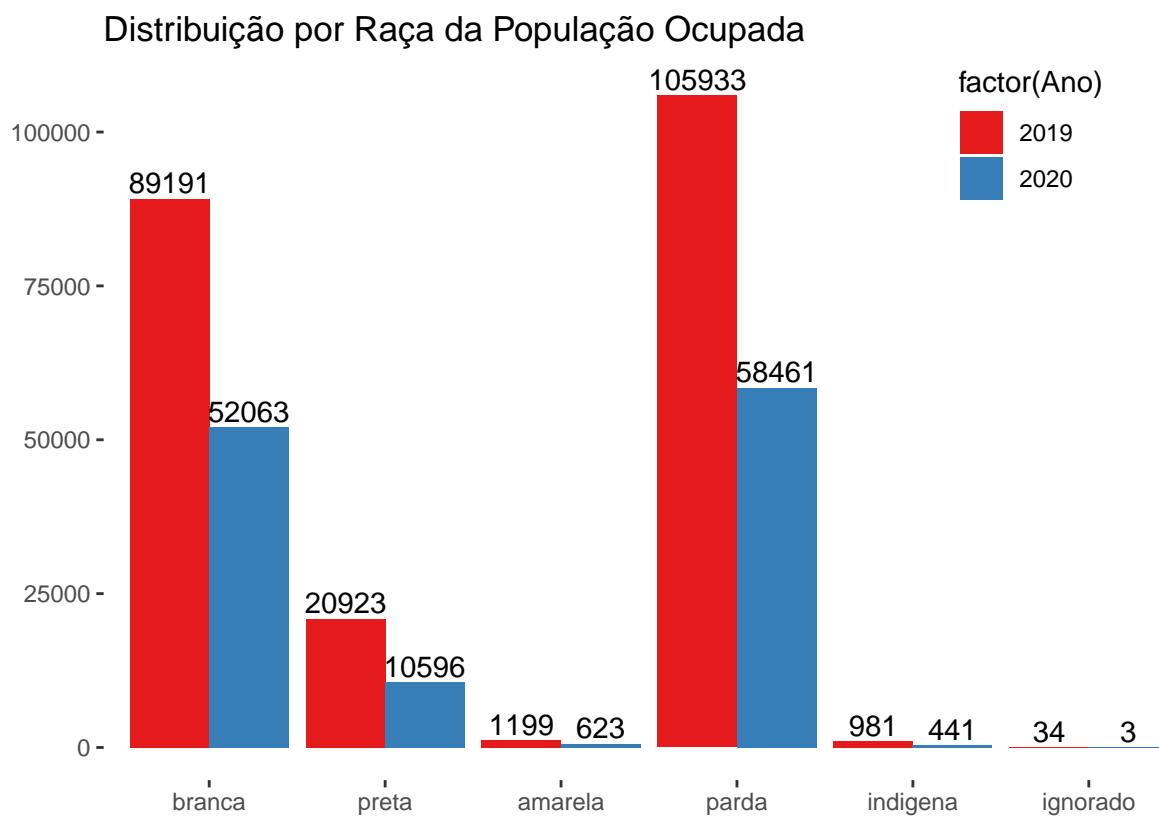


Figure 3: Distribuição por Raça da População Ocupada

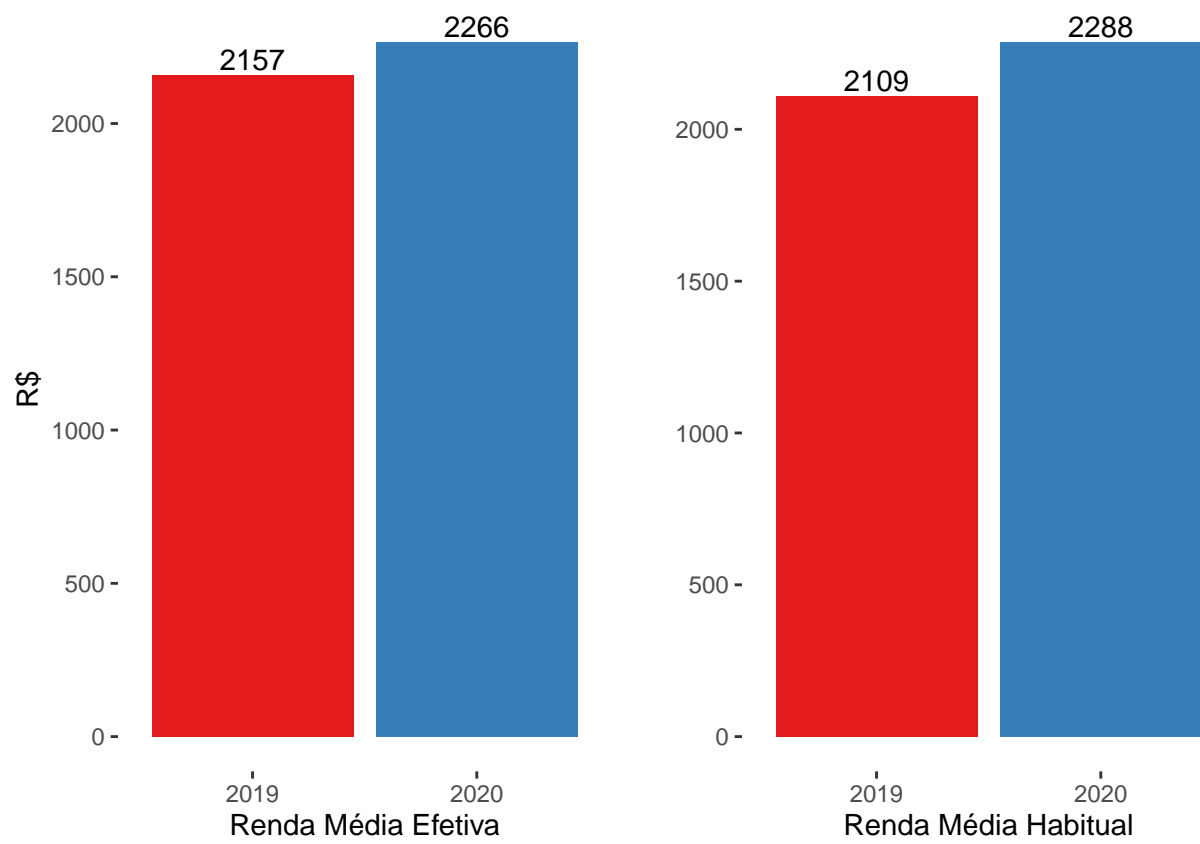


Figure 4: Renda Média dos Ocupados (Habitual e Efetiva)

```
plot4
```

### Análise da Renda Média (Habitual e Efetiva)

```
## TableGrob (1 x 2) "arrange": 2 grobs
##   z      cells   name      grob
## 1 1 (1-1,1-1) arrange gtable[layout]
## 2 2 (1-1,2-2) arrange gtable[layout]

renda <- renda %>% select(-fator)
knitr::kable(renda, caption = "Renda Média Habitual e Efetiva da População Ocupada (2019 e 2020)")
```

Table 6: Renda Média Habitual e Efetiva da População Ocupada (2019 e 2020)

Ano	renda_media_efetiva	renda_media_habitual
2019	2157	2109
2020	2266	2288

## 2. Análise da Renda População Ocupada, com Redução de Jornada e Afastada

Aqui, vamos considerar como regra de bolso que a pessoa está ocupada com jornada reduzida se as horas efetivamente trabalhadas são ao menos 25% menores que as horas habitualmente trabalhadas.

```
# Ocupados com redução de jornada 2019
reducao <- dados %>% filter(VD4002 == 1 & VD4035 < 0.75*VD4031)

# Ocupados, mas temporariamente afastados de seu trabalho.
afastados <- dados %>% filter(Ano == 2020 & VD4002 == 1 & V4005 == 1)

# Ocupados (para comparação)
ocupados <- dados %>% filter(Ano == 2019 & VD4002 == 1)

o19 <- dados %>% filter(Ano == 2019 & VD4002 == 1) %>%
  group_by(Ano, VD4002) %>% summarise("total"=n())
o20 <- dados %>% filter(Ano == 2020 & VD4002 == 1) %>%
  group_by(Ano, VD4002) %>% summarise("total"=n())

or19 <- reducao %>% filter(Ano == 2019 & VD4002 == 1) %>%
  group_by(Ano, VD4002) %>% summarise("total"=n())
or20 <- reducao %>% filter(Ano == 2020 & VD4002 == 1) %>%
  group_by(Ano, VD4002) %>% summarise("total"=n())

oa19 <- afastados %>% filter(Ano == 2019 & VD4002 == 1 & V4005 == 1) %>%
  group_by(Ano, VD4002) %>% summarise("total"=n())

# observa-se que não existem pessoas ocupadas e afastadas no ano de 2019, de acordo com nosso critério
oa19$total

## integer(0)

oa20 <- afastados %>% filter(Ano == 2020 & VD4002 == 1 & V4005 == 1) %>%
  group_by(Ano, VD4002) %>% summarise("total"=n())
```

```

tab5 <-cbind(
rbind("Pop.Ocupada(2019)" = o19$total,
      "Pop.Ocupada (2020)" = o20$total,
      "Var.19_20"= ((o20$total-o19$total)/(o19$total))*100),
rbind("Jornada.Reduzida(2019)" = or19$total,
      "Jornada.Reduzida(2020)" = or20$total,
      "Var.19_20"= ((or20$total-or19$total)/(or19$total))*100),
rbind("Afastados(2019)" = 0,
      "Afastados(2020)" = oa20$total,
      "Var.19_20"= 0)
)
colnames(tab5) <- c("Pop. Ocupada", "Jornada Reduzida", "Afastados")
row.names(tab5) <- c("Ano 2019", "Ano 2020", "Var.19_20")

# visualização dos resultados
knitr::kable(tab5, caption = "Comparativo entre Pop. Ocupada, com Jornada Resumida e Afastada")

```

Table 7: Comparativo entre Pop. Ocupada, com Jornada Resumida e Afastada

	Pop. Ocupada	Jornada Reduzida	Afastados
Ano 2019	218261	10767	0
Ano 2020	122187	9485	3288
Var.19_20	-44	-12	0

```

# filtrando os dados por população ocupada, renda, reduzidos e afastados

renda_ocupados <- dados %>% filter(VD4001==1) %>% group_by(Ano) %>% summarise("renda_efe_ocu"=mean(VD4019,na.rm=TRUE))
"renda_hab_ocu"=mean(VD4019,na.rm=TRUE))

renda_reduzidos <- reducao %>% filter(VD4001==1) %>% group_by(Ano) %>% summarise("renda_efe_red"=mean(VD4019,na.rm=TRUE))
"renda_hab_red"=mean(VD4019,na.rm=TRUE))

renda_afastados <- afastados %>% filter(VD4001==1) %>% group_by(Ano) %>% summarise("renda_efe_afa"=mean(VD4019,na.rm=TRUE))
"renda_hab_afa"=mean(VD4019,na.rm=TRUE))

# observa-se que não existem pessoas ocupadas & afastadas no ano de 2019, de acordo com nosso critério
renda_afastados$Ano; renda_afastados$Ano

```

```
## [1] 2020
```

```
## [1] 2020
```

```

# Filtrando a renda efetiva dos grupos selecionados
efetiva <- rbind(
  rbind(renda_ocupados$renda_efe_ocu),
  rbind(renda_reduzidos$renda_efe_red),
  cbind(0,renda_afastados$renda_efe_afa))

colnames(efetiva) <- c("2019","2020")
row.names(efetiva) <- c("Pop.Ocupada","Jornada Reduzida", "Afastados")

```

```
# visualização dos resultados
```

```
knitr::kable(efetiva, caption = "Renda Média Efetiva das Populações Ocupada, Com Jornada Reduzida e Afastada")
```

Table 8: Renda Média Efetiva das Populações Ocupada, Com Jornada Reduzida e Afastada

	2019	2020
Pop.Ocupada	2157	2266
Jornada Reduzida	2416	2064
Afastados	0	2217

```
# Filtrando a renda habitual dos grupos selecionados
```

```
habitual <- rbind(
  rbind(renda_ocupados$renda_hab_ocu),
  rbind(renda_reduzidos$renda_hab_red),
  cbind(0,renda_afastados$renda_hab_afa))
```

```
colnames(habitual) <- c("2019","2020")
```

```
row.names(habitual) <- c("Pop. Ocupada", "Jornada Reduzida", "Afastados")
```

```
knitr::kable(habitual, caption = "Renda Média Habitual das Populações Ocupada, Com Jornada Reduzida e Afastada")
```

Table 9: Renda Média Habitual das Populações Ocupada, Com Jornada Reduzida e Afastada

	2019	2020
Pop. Ocupada	2109	2288
Jornada Reduzida	2285	2309
Afastados	0	2332