

Coleta, tratamento, visualização e modelagem do TCC

Fábio Rocha | rochaviannafj@gmail.com

2023-02-18

Abstract

O peso ao nascer está diretamente relacionado as chances de doenças na vida adulta e mortalidade ainda na fase neonatal, pós-natal e infância. Compreender, por essa razão, a distribuição do peso dos nascidos vivos e os seus determinantes faz-se necessário já que tais medidas podem sinalizar de forma simples um provável problema de saúde pública que precisa ser enfrentado. Este trabalho surge da necessidade de investigar a relação positiva ou negativa de variáveis (idade da mãe, semanas de gestação, escolaridade da mãe, raça cor da mãe, sexo do bebê e estado civil da mãe), via Método dos Mínimos Quadrados Ordinários, que expliquem o peso dos nascidos vivos em Minas Gerais nos anos 2016-2020, a partir dos dados do DATASUS-SINASC. Os resultados indicam efeito positivo da variável semanas de gestação; as variáveis escolaridade das mães desempenha efeito negativo quando a gestante não tem escolaridade nenhuma; mães pretas tiveram efeito negativo sobre o peso; estado civil apresentou efeito positivo para o caso em que as mães são união estável ou casada; já o sexo do bebê, este apresentou efeito negativo quando o sexo é feminino; consultas pré-natal também desempenham efeito positivo – os resultados aqui reportados indicam ganho médio de 6 gramas a cada consulta adicional. Conclui-se que é necessário (1) ampliação do monitoramento de mães solo – supondo que por estarem solo tem renda menor do que mãe casada ou união estável; (2) acompanhamento das semanas de gestação com o objetivo de identificar gravidez de risco; (3) escolaridade conta positivamente já que mães mais escolarizadas em geral tem melhores rendimentos; (4) Sugere-se, por fim, estudos complementares que considere variáveis do contexto político e da pandemia do Covid-19.

Pacotes usados no estudo

```
library(microdatasus)
library(tidyverse)
library(stargazer)
library(lubridate)
```

Fazendo o download dos dados do FTP do Datasus

```
#nasc <- fetch_datasus(year_start = 2016,
#                       year_end = 2020,
#                       uf = "MG",
#                       information_system = "SINASC")
```

Fazendo um tratamento prévio dos dados com função própria do pacote Microdatasus

```
#nascimento <- process_sinasc(nasc, municipality_data = FALSE)
```

Selecionando as variáveis que serão utilizadas em meu estudo

```
#write.csv(nascimento, 'C:/Users/Usuario/Desktop/22_2/ATV600/Técnicas de Pesquisa em Economia II/dados.csv',  
#          row.names = FALSE)  
  
nascimento = read.csv("dados.csv")  
  
selecao_nascimento <- nascimento %>%  
  select(DTNASC,  
         PESO,  
         IDADEMAE,  
         CONSPRENAT,  
         ESCMAE,  
         RACACORMAE,  
         SEXO,  
         ESTCIVMAE,  
         GESTACAO)
```

Criação de variáveis dummy para variável raça/cor e uma variável dummy para pandemia¹

```
selecao_nascimento <- selecao_nascimento %>%  
  mutate(Parda = ifelse(selecao_nascimento$RACACORMAE == "Parda", 1, 0)) %>%  
  mutate(Indigena = ifelse(selecao_nascimento$RACACORMAE == "Indígena", 1, 0)) %>%  
  mutate(Branca = ifelse(selecao_nascimento$RACACORMAE == "Branca", 1, 0)) %>%  
  mutate(Preta = ifelse(selecao_nascimento$RACACORMAE == "Preta", 1, 0)) %>%  
  mutate(pandemia = ifelse(selecao_nascimento$DTNASC >= "2020-11-01", 1, 0))
```

Criação de variável negro e não negro

```
selecao_nascimento <- selecao_nascimento %>%  
  mutate(negro = if_else(selecao_nascimento$RACACORMAE == "Parda"|selecao_nascimento$RACACORMAE == "Preto", 1, 0))
```

Criação de variável dummy para variável negro e não negro

```
selecao_nascimento <- selecao_nascimento %>%  
  mutate(nao_negro = ifelse(selecao_nascimento$negro == "Não negro", 1, 0))
```

Criação de variável dummy para variável gênero

```
selecao_nascimento <- selecao_nascimento %>%  
  mutate(feminino = ifelse(selecao_nascimento$SEXO == "Feminino", 1, 0))
```

¹Para estudos futuros

Conversão das variáveis idade e peso para numéricas

```
selecao_nascimento$PESO <- as.integer(selecao_nascimento$PESO)
selecao_nascimento$IDADEMAE <- as.integer(selecao_nascimento$IDADEMAE)
```

Identificando quantidade de mães com idade superior a 60 anos

```
selecao_nascimento %>%
  count(IDADEMAE > 60)
```

```
##   IDADEMAE > 60      n
## 1          FALSE 607996
## 2           TRUE      3
```

Foram identificados 9 mães com idade superior a 60 e 127 com idade superior a 50

```
selecao_nascimento %>%
  count(IDADEMAE > 50)
```

```
##   IDADEMAE > 50      n
## 1          FALSE 607927
## 2           TRUE      72
```

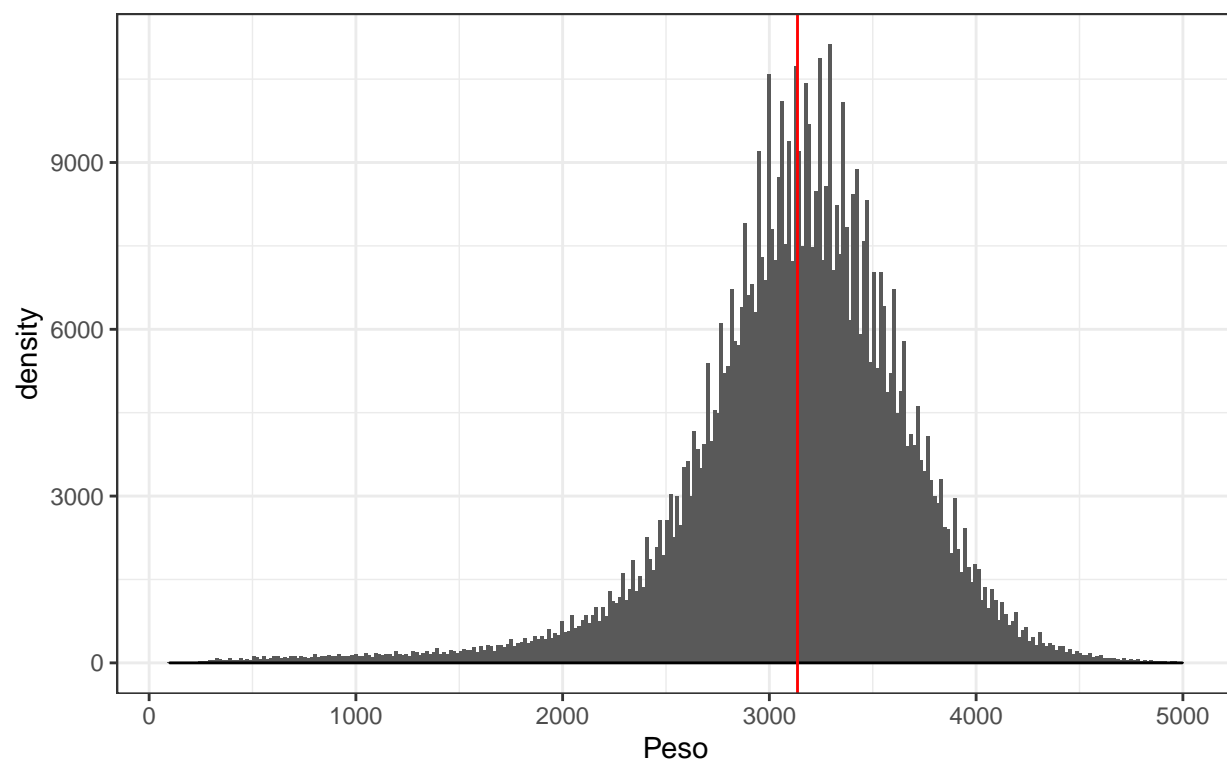
Filtrando apenas bebês que nasceram com peso inferior a 5kg

```
selecao_nascimento <- selecao_nascimento %>%
  filter(PESO < 5000)
```

Com este filtro, foram removidos 620 observações. Ou seja, 620 bebês tinha peso superior a 5 kg.

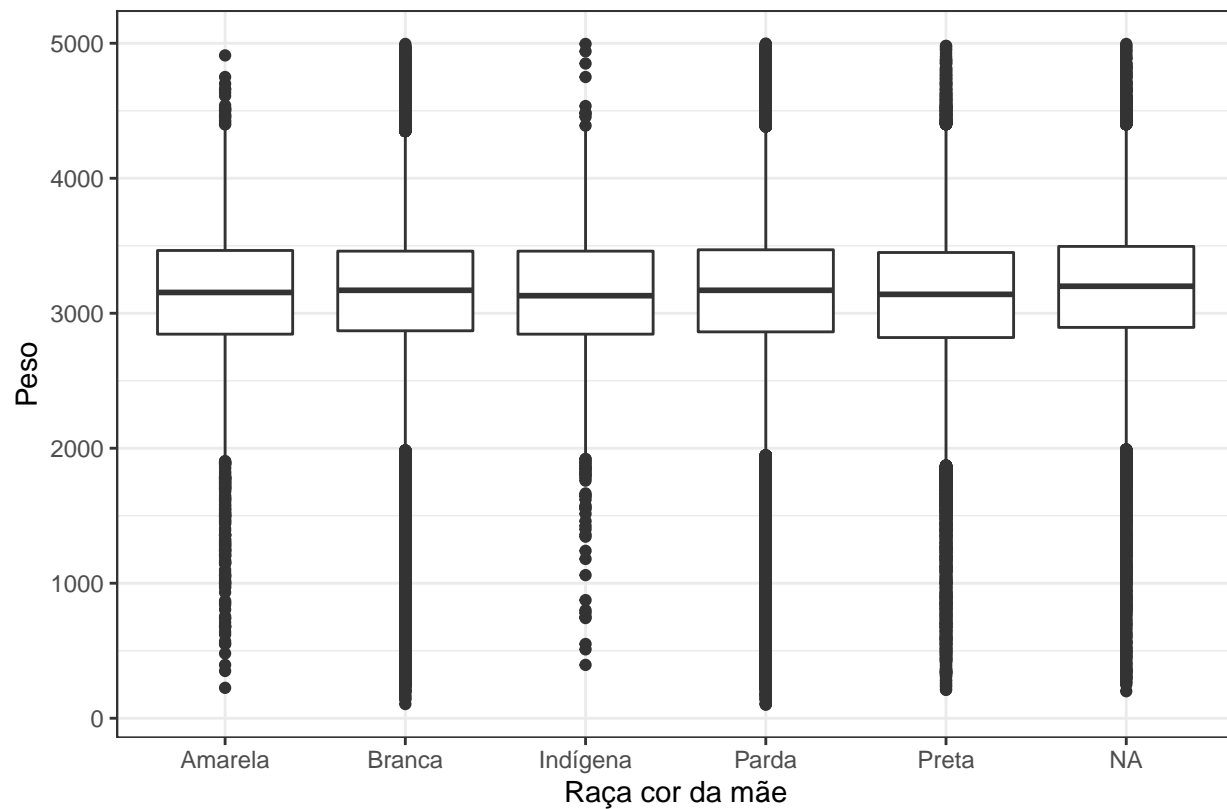
Alguns gráficos - Descritiva do Estudo

```
ggplot(selecao_nascimento, aes(x=PESO))+
  geom_histogram(bins = 300)+
  geom_density()+
  geom_vline(xintercept=3136, colour='red')+
  theme_bw()+
  labs(title = "",
       x="Peso",
       caption="Fonte: DATASUS - SINASC")
```



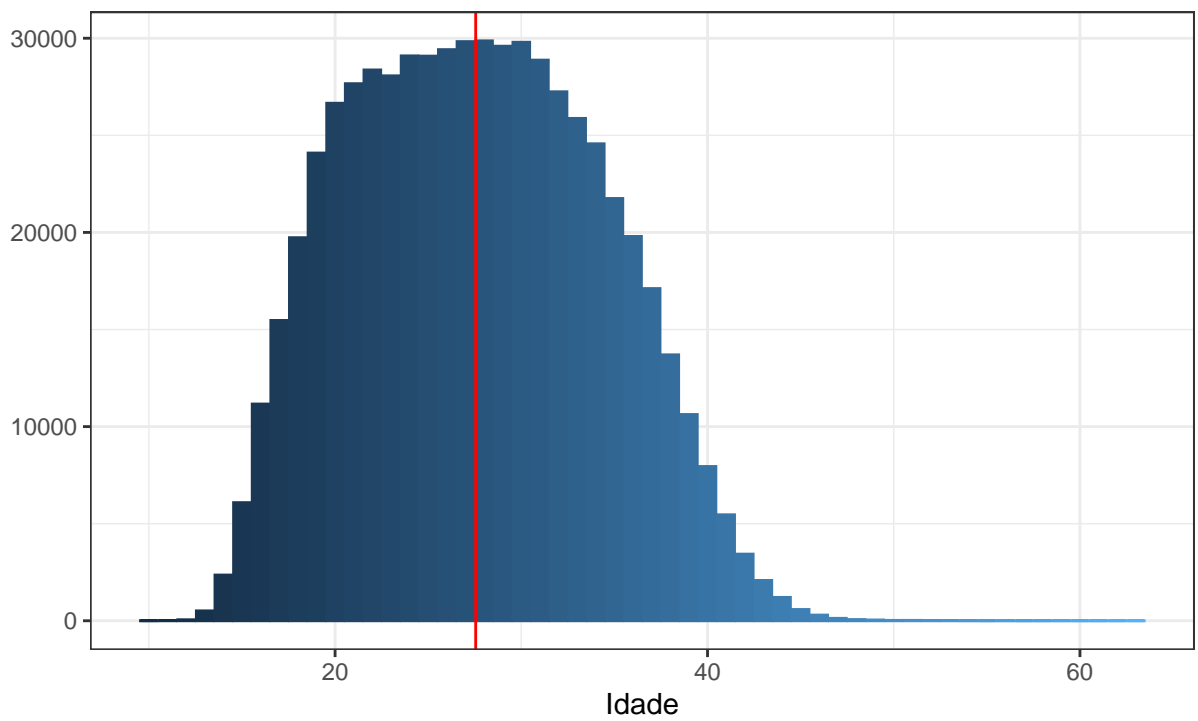
Fonte: DATASUS – SINASC

```
ggplot(selecao_nascimento, aes(y=PESO, x=RACACORMAE))+  
  geom_boxplot()+  
  theme_bw()+  
  labs(x="Raça cor da mãe",  
        y="Peso",  
        caption = "Fonte: DATASUS - SINASC")
```



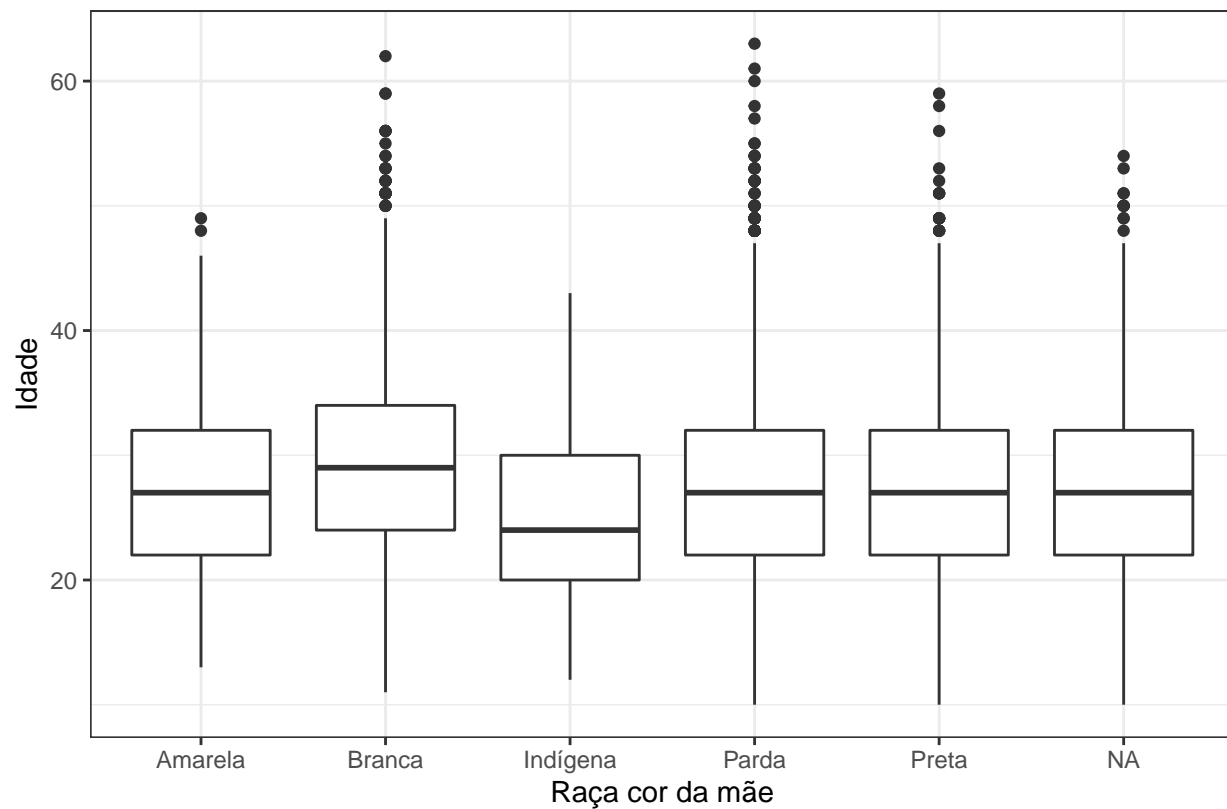
Fonte: DATASUS – SINASC

```
selecao_nascimento %>%
  count(IDADEMAE) %>%
  ggplot(aes(x=IDADEMAE,
             y=n,
             fill=IDADEMAE,
             color=IDADEMAE))+
  geom_bar(stat = "identity")+
  geom_vline(xintercept=27.55, colour='red')+
  labs(title = "",
       subtitle = "",
       x = "Idade",
       y = "",
       caption = "Fonte: DATASUS - SINASC")+
  theme_bw()+
  theme(legend.position = "none")
```



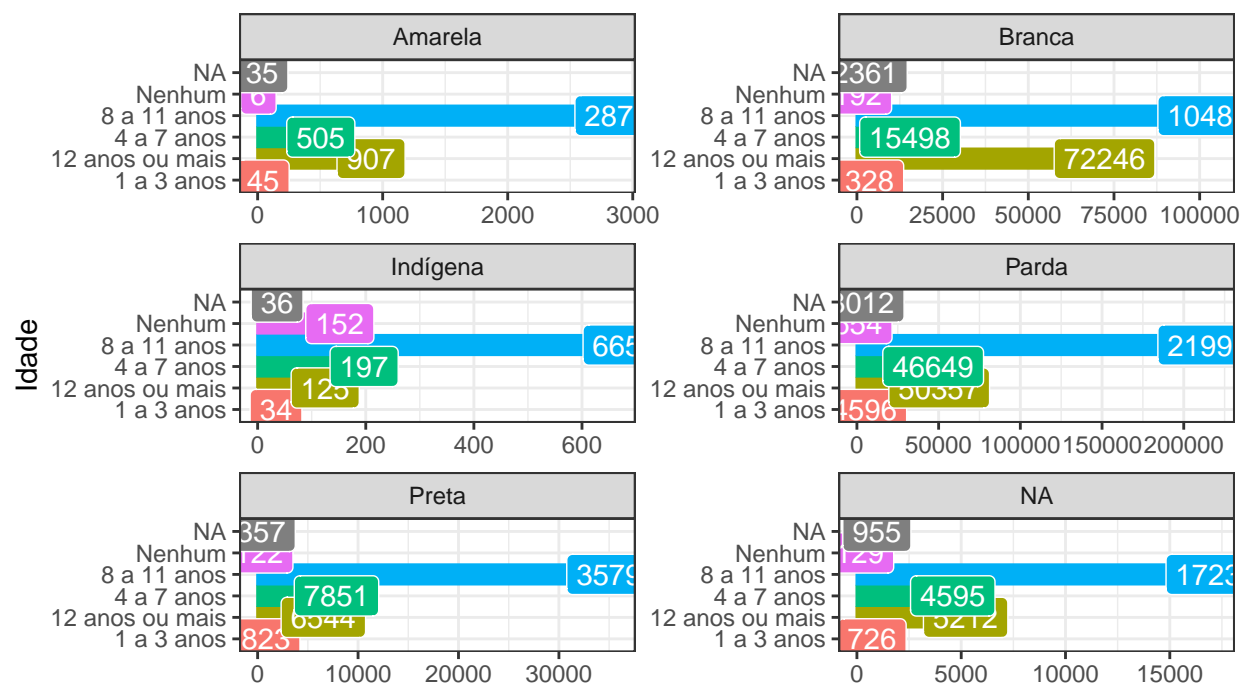
Fonte: DATASUS – SINASC

```
ggplot(selecao_nascimento, aes(y=IDADEMAE, x=RACACORMAE))+  
  geom_boxplot()+  
  theme_bw()+  
  labs(x="Raça cor da mãe",  
        y="Idade",  
        caption = "Fonte: DATASUS - SINASC")
```

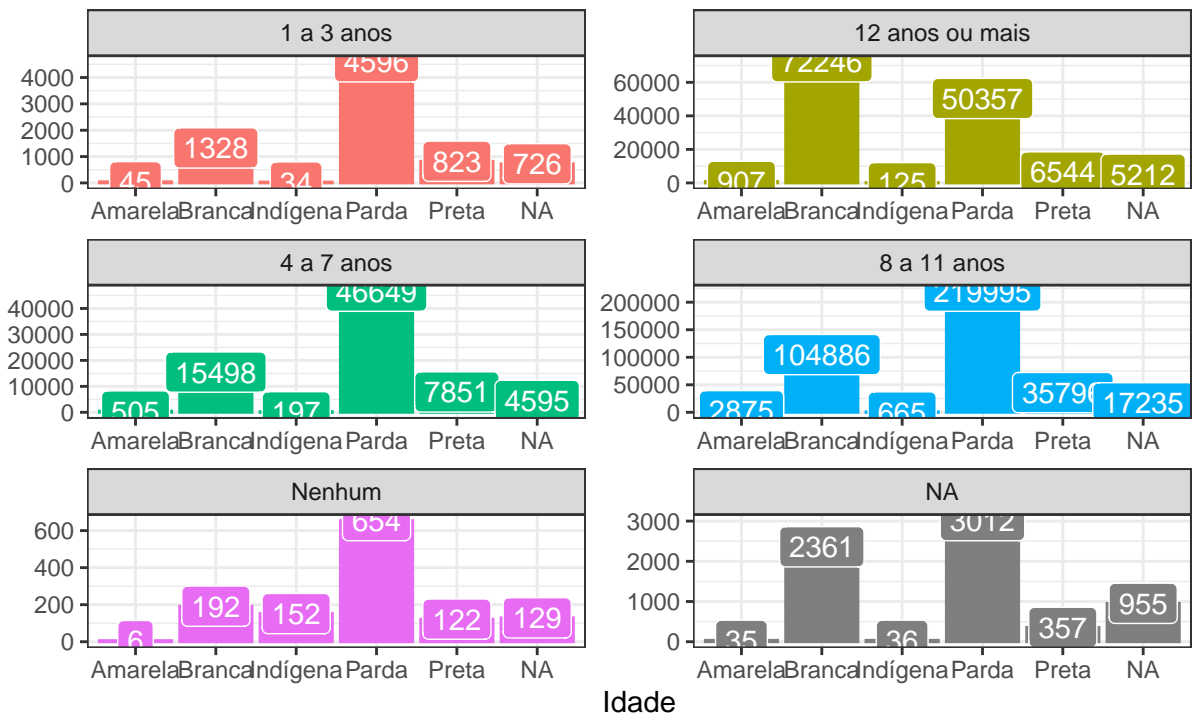


Fonte: DATASUS – SINASC

```
selecao_nascimento %>%
  count(ESCMAC, RACACORMAC) %>%
  ggplot(aes(x=ESCMAC,
             y=n,
             fill=ESCMAC,
             color=ESCMAC,
             label=n))+
  geom_bar(stat = "identity")+
  geom_label(color="white")+
  facet_wrap(~RACACORMAC, scales = "free", ncol = 2 )+
  coord_flip()+
  labs(title = "",
       subtitle = "",
       x = "Idade",
       y = "",
       caption = "")+
  theme_bw()+
  theme(legend.position = "none")
```

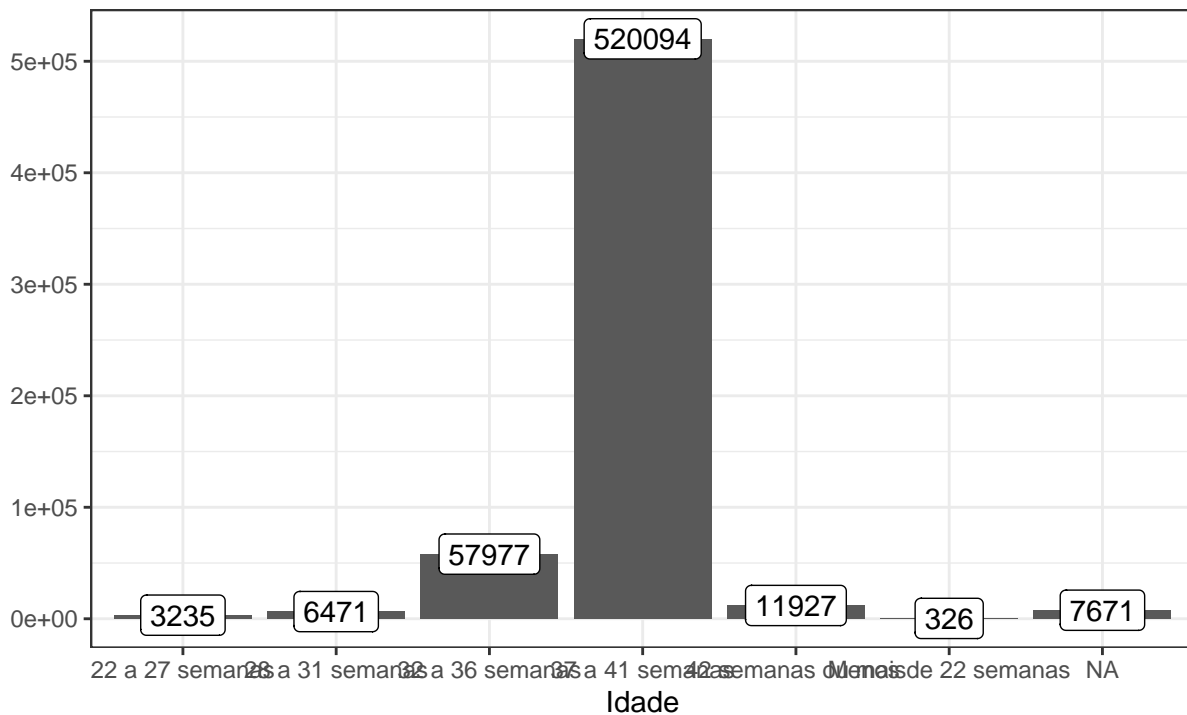


```
selecao_nascimento %>%
  count(ESCMAC, RACACORMAE) %>%
  ggplot(aes(x=RACACORMAE,
             y=n,
             fill=ESCMAC,
             color=ESCMAC,
             label=n))+
  geom_bar(stat = "identity")+
  geom_label(color="white")+
  facet_wrap(~ESCMAC, scales = "free", ncol = 2 )+
  labs(title = "",
       subtitle = "",
       x = "Idade",
       y = "",
       caption = "Fonte: DATASUS - SINASC")+
  theme_bw()+
  theme(legend.position = "none")
```

Fonte: DATASUS – SINASC

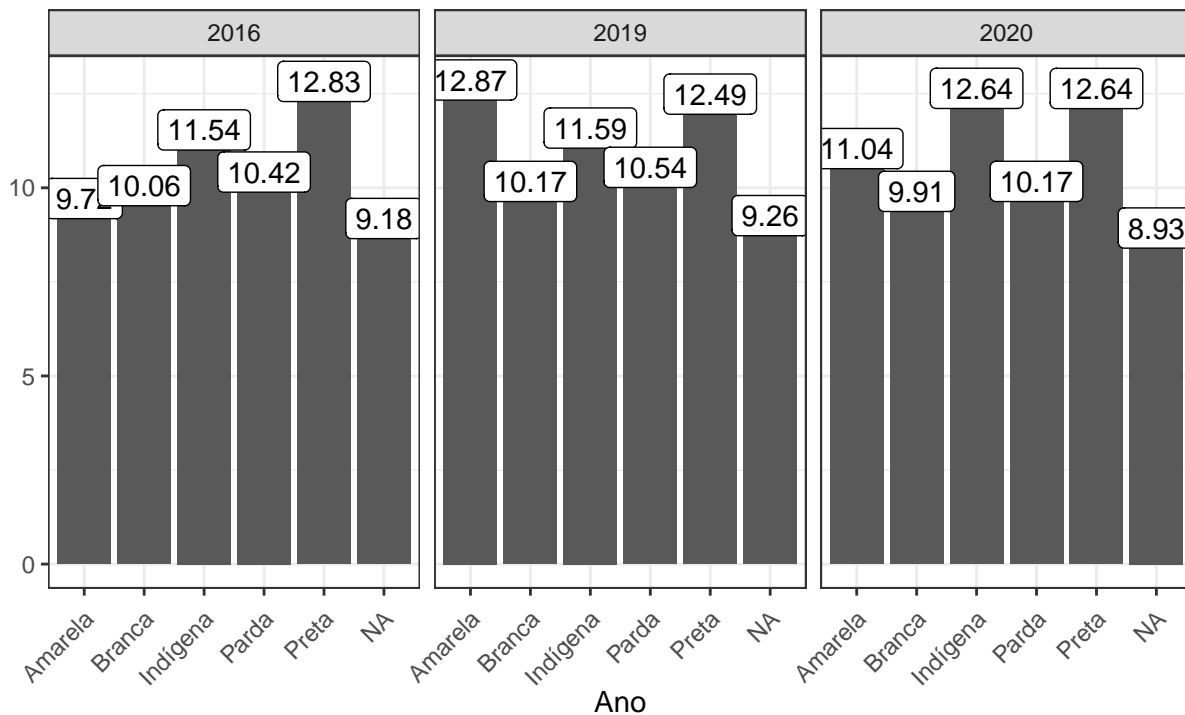
```
selecao_nascimento %>%
  count(GESTACAO) %>%
  ggplot(aes(x=GESTACAO,
             y=n,
             label=n))+
  geom_bar(stat = "identity")+
  geom_label(color="black")+
  #coord_flip()+
  labs(title = "",
        subtitle = "",
        x = "Idade",
        y = "",
        caption = "Fonte: DATASUS - SINASC")+
  theme_bw()+
  theme(legend.position = "none")
```



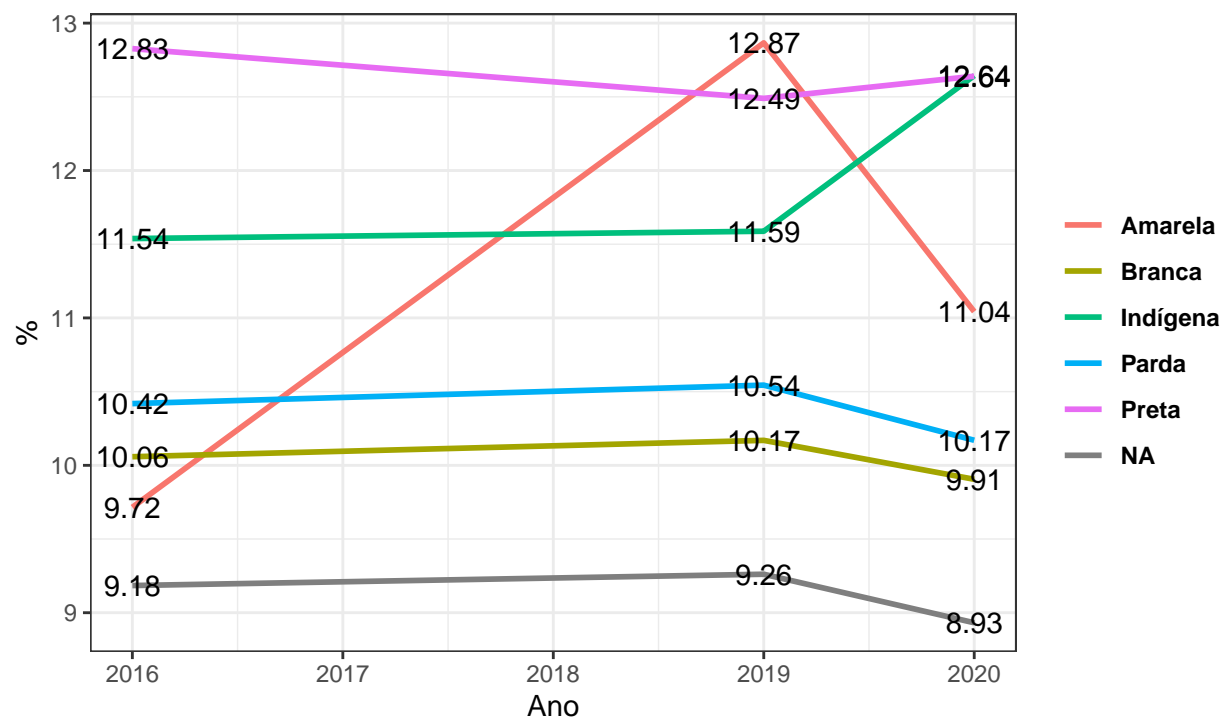
```
bpn <- selecao_nascimento %>%
  mutate(bpn = ifelse(PESO < 2500, "baixo peso", "não baixo peso"))
```

```
bpn_grafico_prop <- bpn %>%  
  count(bpn, ano, RACACORMAE)
```

```
ggplot(bpn_grafico_prop, aes(x=RACACORMAE, y=proporcao))+
  geom_col()+
  theme_bw()+
  facet_wrap(~as.character(ano), nrow = 2, ncol = 3)+
  theme(axis.text.x = element_text(angle = 45, hjust = 1, vjust = 0.9))+
  geom_label(aes(label=round(proporcao,2)))+
  labs(x='Ano',
       y='',
       title="",
       subtitle = "",
       caption='')
```



```
ggplot(bpn_grafico_prop, aes(x=ano, y=proporcao))+
  geom_line(aes(color=RACACORMAE),size=1)+
  theme_bw()+
  theme(#legend.position = "right",
        legend.title = element_blank(),
        plot.title = ggplot2::element_text(face = "bold"),
        plot.subtitle = ggplot2::element_text(face = "italic"),
        legend.text = ggplot2::element_text(face = "bold"))+
  geom_text(aes(label=round(proporcao,2)), size=3.8)+
  labs(x='Ano',
        y='%',
        title="",
        subtitle = "",
        caption='Fonte: DATASUS - SINASC')
```



Fonte: DATASUS – SINASC

Desvio padrão do peso e idade

```
sd(selecao_nascimento$PESO, na.rm = TRUE)
```

```
[1] 547.2024
```

```
sd(selecao_nascimento$IDADEMAE, na.rm = TRUE)
```

```
[1] 6.693453
```

Primeiro Modelo

```
modelo <- lm(PESO ~ IDADEMAE + as.integer(CONSPRENAT)+ ESCMAE + RACACORMAE + SEXO + ESTCIVMAE + GESTACAO)
```

```
modelo_final <- lm(PESO ~ IDADEMAE +
  as.integer(CONSPRENAT) +
  ESCMAE +
  RACACORMAE +
  feminino +
  ESTCIVMAE +
  GESTACAO, selecao_nascimento)
```

Tratamento dos resíduos - Erros robustos de White

```
library(car)
library(AER)

homoce.final <- hccm(modelo_final, type=c('hc1'))
modelo.final<- coeftest(modelo_final, homoce.final)
modelo.final

##
## t test of coefficients:
##
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)      971.91117    17.54775  55.3867 < 2.2e-16 ***
## IDADEMAE           1.96037     0.10500  18.6709 < 2.2e-16 ***
## as.integer(CONSPRENAT)  6.56669     0.21193  30.9857 < 2.2e-16 ***
## ESCMAE12 anos ou mais  0.73825     6.48342   0.1139  0.90934
## ESCMAE4 a 7 anos     15.13662     6.60216   2.2927  0.02187 *
## ESCMAE8 a 11 anos    18.92112     6.39727   2.9577  0.00310 **
## ESCMAENenhum       -32.58719    17.50645  -1.8614  0.06268 .
## RACACORMAEBranca      5.55659     7.08973   0.7838  0.43319
## RACACORMAEIndígena    35.20324    16.32575   2.1563  0.03106 *
## RACACORMAEParda      16.14714     7.05958   2.2873  0.02218 *
## RACACORMAEPreta      -7.64783     7.31805  -1.0451  0.29599
## feminino           -116.90681     1.21684 -96.0743 < 2.2e-16 ***
## ESTCIVMAESeparada judicialmente  4.08926     4.57078   0.8947  0.37097
## ESTCIVMAESolteira    -19.89217     1.46068 -13.6184 < 2.2e-16 ***
## ESTCIVMAEUnião consensual  11.14268     2.07191   5.3780 7.535e-08 ***
## ESTCIVMAEViúva      -19.42628    12.44310  -1.5612  0.11848
## GESTACAO28 a 31 semanas  603.31724    17.13293  35.2139 < 2.2e-16 ***
## GESTACAO32 a 36 semanas 1544.24993    14.48070 106.6419 < 2.2e-16 ***
## GESTACAO37 a 41 semanas 2170.35712    14.29428 151.8340 < 2.2e-16 ***
## GESTACAO42 semanas ou mais 2234.43949    14.89857 149.9768 < 2.2e-16 ***
## GESTACAOMenos de 22 semanas  94.41391    70.54874   1.3383  0.18081
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Modelo com variavel pandemia

```
modelo_pandemia <- lm(PESO ~ IDADEMAE +
  as.integer(CONSPRENAT) +
  pandemia +
  ESCMAE +
  RACACORMAE +
  feminino +
  ESTCIVMAE +
  GESTACAO, selecao_nascimento)
```

```
stargazer::stargazer(modelo, modelo_final, modelo_pandemia,
  font.size = 'tiny',
  header = FALSE,
  title="Resultados")
```

Table 1: Resultados

	<i>Dependent variable:</i>		
	PESO		
	(1)	(2)	(3)
IDADEMAE	1.960*** (0.102)	1.960*** (0.102)	1.960*** (0.102)
as.integer(CONSPRENAT)	6.567*** (0.184)	6.567*** (0.184)	6.567*** (0.184)
pandemia			
ESCMAE12 anos ou mais	0.738 (5.854)	0.738 (5.854)	0.738 (5.854)
ESCMAE4 a 7 anos	15.137** (5.945)	15.137** (5.945)	15.137** (5.945)
ESCMAE8 a 11 anos	18.921*** (5.749)	18.921*** (5.749)	18.921*** (5.749)
ESCMAE Nenhum	-32.587** (15.436)	-32.587** (15.436)	-32.587** (15.436)
RACACORMAE Branca	5.557 (7.074)	5.557 (7.074)	5.557 (7.074)
RACACORMAE Indígena	35.203** (15.460)	35.203** (15.460)	35.203** (15.460)
RACACORMAE Parda	16.147** (7.039)	16.147** (7.039)	16.147** (7.039)
RACACORMAE Preta	-7.648 (7.283)	-7.648 (7.283)	-7.648 (7.283)
SEXOMasculino	116.907*** (1.218)		
feminino		-116.907*** (1.218)	-116.907*** (1.218)
ESTCIVMAE Separada judicialmente	4.089 (4.525)	4.089 (4.525)	4.089 (4.525)
ESTCIVMAE Solteira	-19.892*** (1.450)	-19.892*** (1.450)	-19.892*** (1.450)
ESTCIVMAE União consensual	11.143*** (2.035)	11.143*** (2.035)	11.143*** (2.035)
ESTCIVMAE Viúva	-19.426 (12.460)	-19.426 (12.460)	-19.426 (12.460)
GESTACAO 28 a 31 semanas	603.317*** (10.236)	603.317*** (10.236)	603.317*** (10.236)
GESTACAO 32 a 36 semanas	1,544.250*** (8.617)	1,544.250*** (8.617)	1,544.250*** (8.617)
GESTACAO 37 a 41 semanas	2,170.357*** (8.430)	2,170.357*** (8.430)	2,170.357*** (8.430)
GESTACAO 42 semanas ou mais	2,234.439*** (9.470)	2,234.439*** (9.470)	2,234.439*** (9.470)
GESTACAO Menos de 22 semanas	94.414*** (28.237)	94.414*** (28.237)	94.414*** (28.237)
Constant	855.004*** (12.833)	971.911*** (12.829)	971.911*** (12.829)
Observations	561,289	561,289	561,289
R ²	0.301	0.301	0.301
Adjusted R ²	0.301	0.301	0.301
Residual Std. Error (df = 561268)	456.006	456.006	456.006
F Statistic (df = 20; 561268)	12,069.490***	12,069.490***	12,069.490***

Note:

* p<0.1; ** p<0.05; *** p<0.01