

[31]	"kidstotal"	"wannakids"	"paidjob"
[34]	"assoc"	"levelasOUTDATED"	"levelassoc"
[37]	"familysupport"	"partnersupport"	"partnchangcitsupport"
[40]	"candwchance"	"thinkcand"	"sex"
[43]	"ageinterval"	"educalevel"	"incomeinterval"
[46]	"religion"	"pesos"	

É importante visualizar a base de dados diretamente, bem como o resumo da importação dos dados (summary()), para garantir uma leitura correta das variáveis e evitar perda de dados. Ao fazer isso, identificamos a falha na leitura de "polbr2", o que nos fez retornar e corrigir onde possível, sabendo que variáveis com muitos dados faltantes devem ser evitadas no modelo, para não conduzirem a conclusões enviesadas.

Detalhamento da amostra

Nessa seção, segue-se visualização dos dados em análise exploratória preliminar, para identificar os tipos de variáveis e sua distribuição.

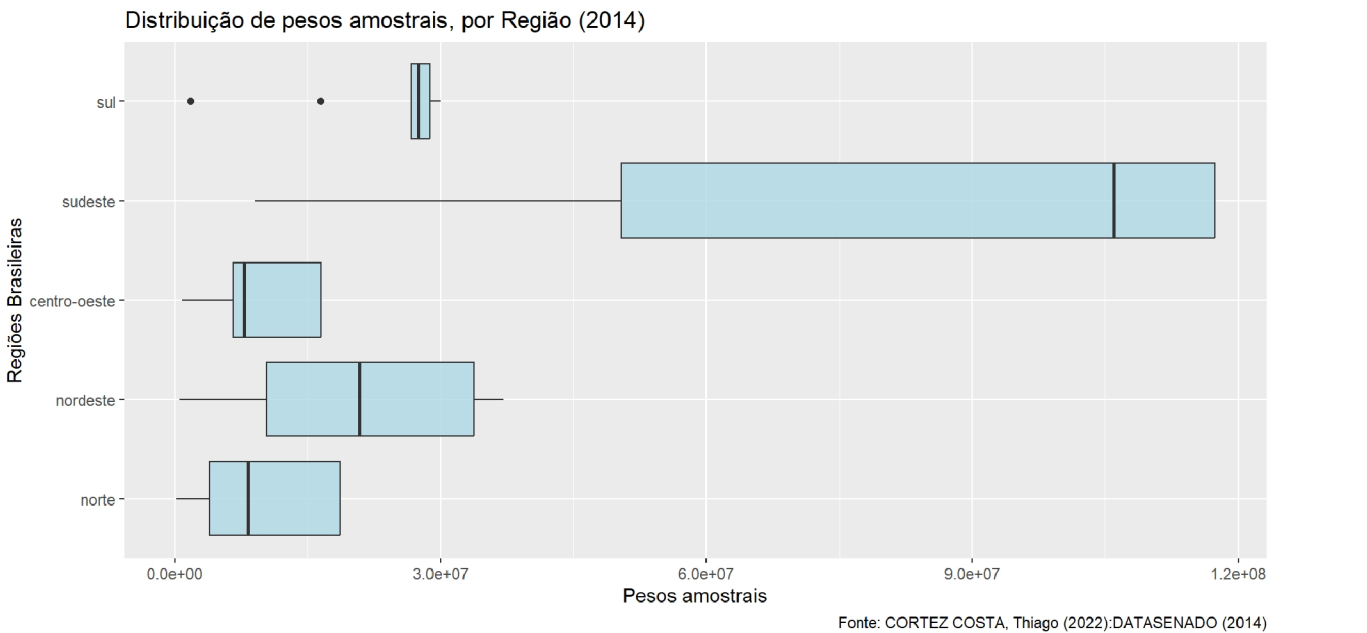
A (fig_pesosamostraisdist?) mostra a distribuição dos pesos amostrais por região. É possível observar como a coleta de dados foi regionalizada, para incluir observações em regiões menos populosas e aumentar a eficiência das estimativas.

(fig_pesosamostraisdist?) - Distribuição dos pesos amostrais, por região (DATASENADO, 2014)

▼ Código

```
```{r}
#| column: screen-inset-shaded
#| layout-nrow: 6
#| label: fig_pesosamostraisdist
#| fig-cap: Distribuição dos pesos amostrais, por região do Brasil
#| fig-subcap: FONTE: CORTEZ COSTA, Thiago. 2022

ggplot(
 data = polbr,
 mapping =
 aes(x = regioao,
 y = pesos
)
) +
 geom_boxplot(fill = c("light blue"), alpha = 0.80) +
 labs(x = "Regiões Brasileiras",
 y = "Pesos amostrais",
 title =
 "Distribuição de pesos amostrais, por Região (2014)",
 caption=
 "Fonte: CORTEZ COSTA, Thiago (2022):DATASENADO (2014)") +
 coord_flip()
```
```



▼ Código

```
```{r}
polbr %>% count(ageinterval) %>% mutate (pct = n / sum (n)) %>% ggplot(mapping = aes(x = ageinterval, y = pct)) + geom_col(fill="light t
```
```