

Laboratório 2

Camille Menezes e Michel Miler

03 de abril de 2023

Parabéns, o texto tem boa fluidez.

1 Medidas morfológicas dos gambás

1.1 Descrevendo o comportamento de cada uma das variáveis

A Tabela 1, indica que as duas variáveis são aproximadamente simétricas, pois as médias são próximas as medianas e as modas, embora a amplitude dos primeiros dois quantis do comprimento total seja maior. Olhando para a variância, as observações acerca da largura do crânio estão mais concentradas ao redor da média.

Tabela 1: Sumarização das variáveis largura do crânio (*skullw*) e comprimento total (*totlngth*) de gambás das montanhas capturados na Inglaterra

	<i>skullw</i>	<i>totlngth</i>
Mínimo	50,00	75,00
1º quartil	54,98	84,00
Mediana	56,35	88,00
Média	56,88	87,09
Moda	57,60	89,00
3º quartil	58,10	90,00
Máximo	68,60	96,50
Variância	9,693	18,58

Através da Figura 1, é possível notar que há pontos atípicos na variável que corresponde a largura do crânio dos animais.

Em relação a largura do crânio, as medianas entre machos e fêmeas são próximas (Figura 2), mas a variabilidade da dos machos é maior, principalmente acima da mediana. Há valores atípicos para para a largura do crânio em ambos os sexos. Já para o comprimento total, a mediana das fêmeas é maior que a dos machos. A maior variabilidade para as fêmeas está entre primeiro quartil e a mediana e, para os machos, está entre a mediana e o terceiro quartil.

Figura 1: *Boxplots* da largura do crânio (*skullw*) e comprimento total (*totlngth*) de gambás das montanhas capturados na Inglaterra

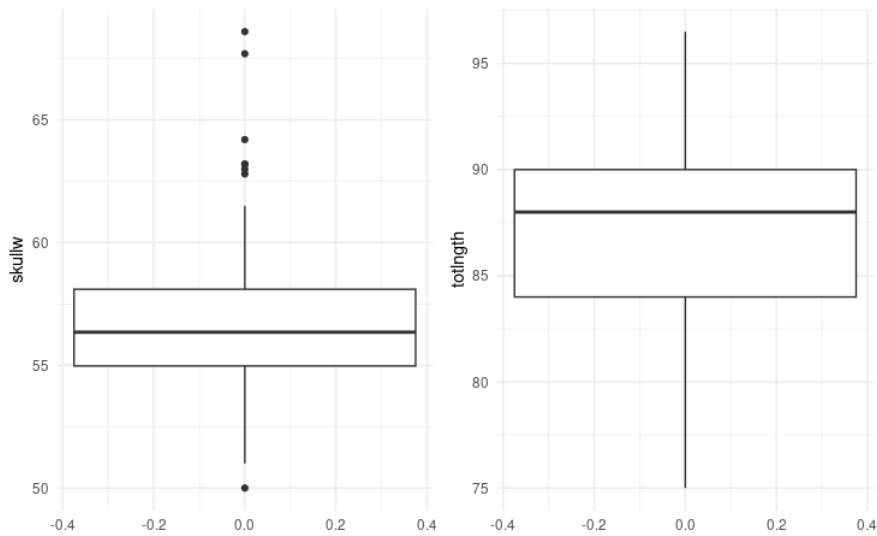
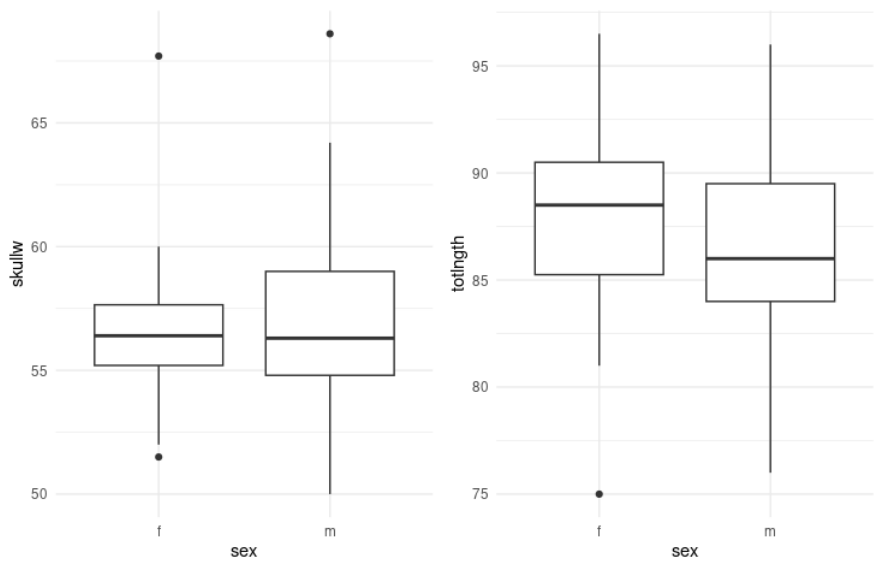


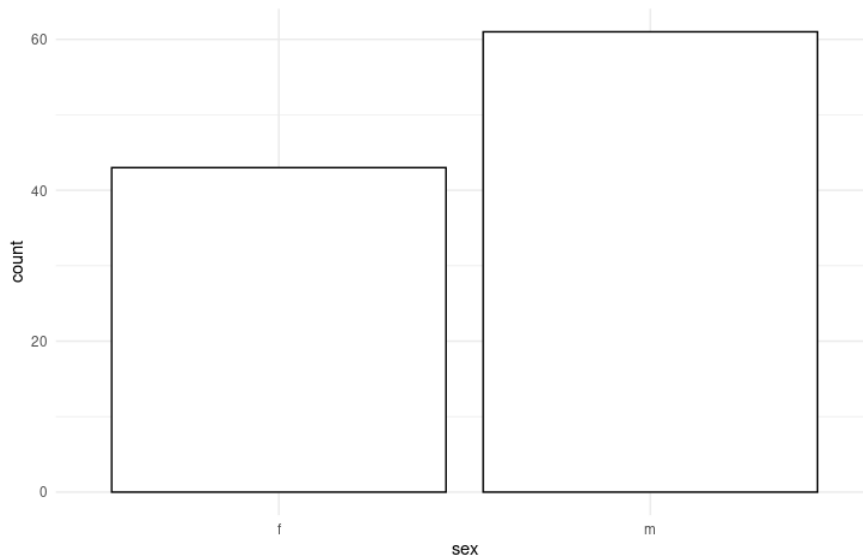
Figura 2: *Boxplots*, de acordo com o sexo, da largura do crânio (*skullw*) e comprimento total (*totlngth*) de gambás das montanhas capturados na Inglaterra



1.2 Representando graficamente a distribuição do sexo

Os machos compõem 58,65% da população enquanto as fêmeas representam 44,35%. Apesar de haver mais gambás machos que fêmeas, essa diferença não é tão acentuada, o que significa que é possível comparar os dois sexos sem prejuízos.

Figura 3: Distribuição do sexo de gambás das montanhas capturados na Inglaterra



1.3 Relações entre as variáveis morfológicas

Através da Figura 4, é possível notar que as variáveis que representam a largura do crânio e o comprimento total apresentam um comportamento que se assemelha a uma distribuição unimodal. O gráfico de dispersão (Figura 5) entre essas duas variáveis, evidencia que existe uma relação linear positiva entre elas. Além disso, o coeficiente de correlação de Pearson é 0,526, indicando que há uma correlação positiva moderada entre essas variáveis.

Figura 4: Histograma da largura do crânio (*skullw*) e do comprimento total (*totlngth*) de gambás das montanhas capturados na Inglaterra

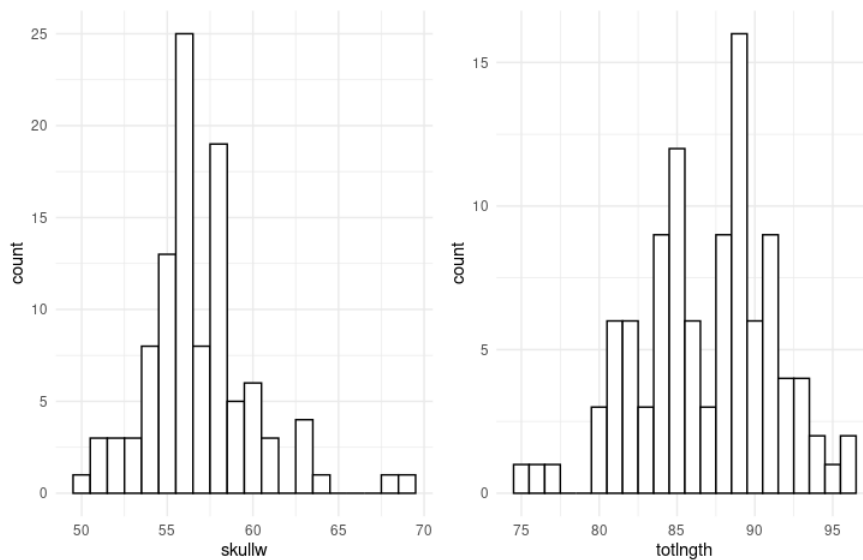
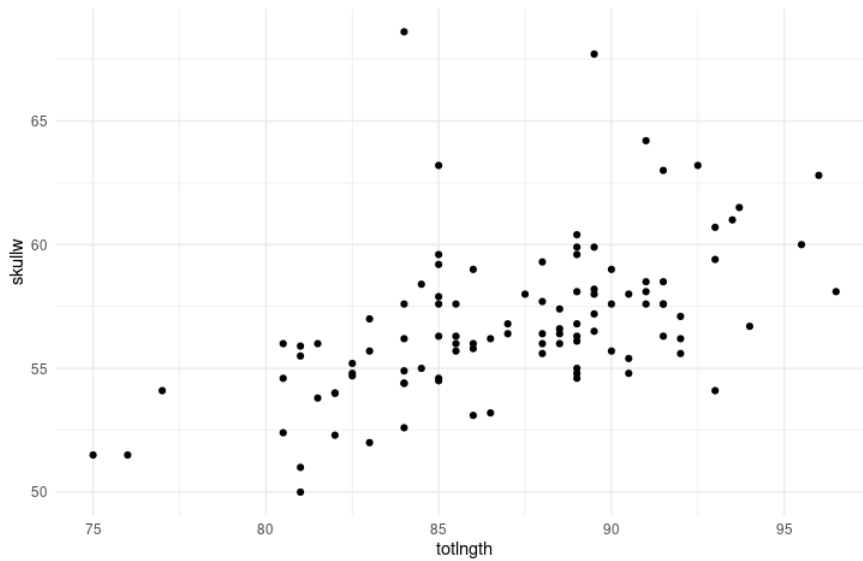


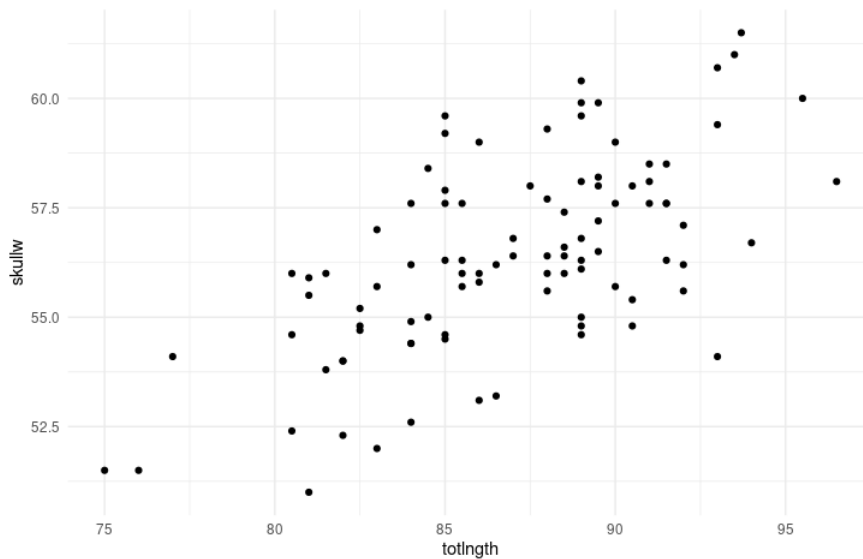
Figura 5: Gráfico de dispersão da largura do crânio (*skullw*) e comprimento total (*totlngth*) de gambás das montanhas capturados na Inglaterra



1.4 Refazendo as análises do item anterior sem a presença de pontos atípicos

Através da Figura 1, é possível observar que há a presença de valores atípicos na variáveis que representa a largura do crânio e, por isso, eles foram retirados e as análises foram refeitas.

Figura 6: Gráfico de dispersão da largura do crânio (*skullw*) e comprimento total (*totlngth*) de gambás das montanhas capturados na Inglaterra



Sem os pontos atípicos, a relação linear positiva aparenta ser ainda mais forte, isso pode ser observado no gráfico de dispersão (Figura 6). O coeficiente de correlação de Pearson aumentou para 0,626, reiterando o que foi percebido através do gráfico.

1.5 Ajuste do modelo de regressão linear

Devido ao o que foi discutido anteriormente, foi possível concluir que parece ser adequado ajustar um modelo de regressão linear para esses dados.

As estimativas para os parâmetros foram $\hat{\beta}_0 = 19,422$ e $\hat{\beta}_1 = 1,198$. O valor estimado para o β_1 nos indica que a relação linear é positiva, pois este valor é maior que 0 . Já o β_0 estimado não tem interpretabilidade, pois o animal não pode ter a largura do crânio igual a 0 .

Figura 7: Gráfico de dispersão e curva de regressão linear ajustada aos dados de gambás das montanhas capturados na Inglaterra

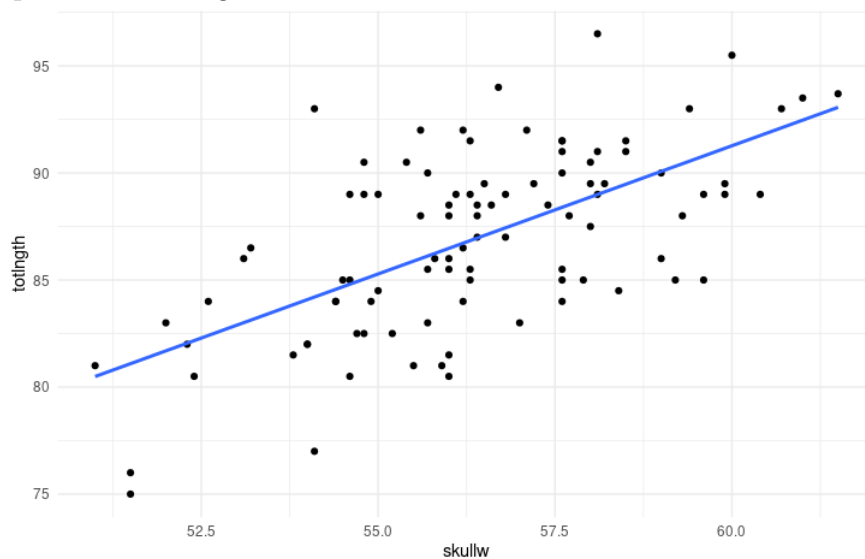
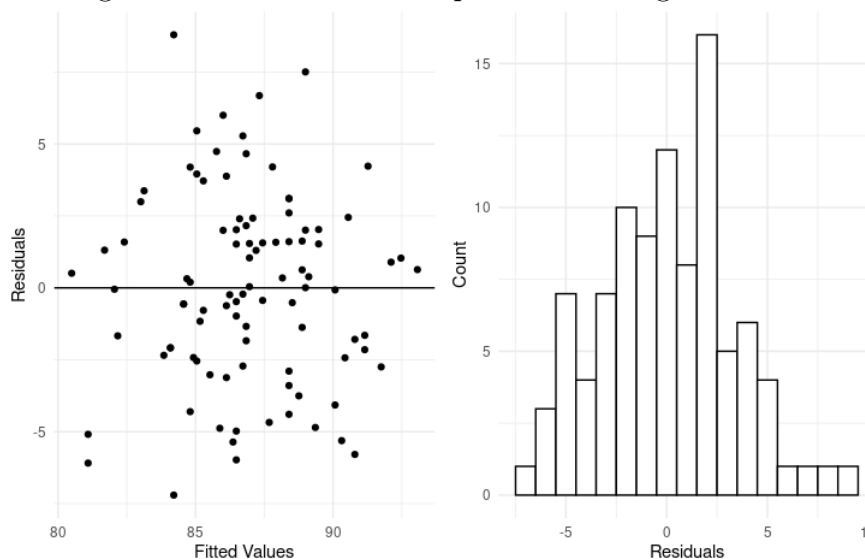


Figura 8: Gráfico dos resíduos versus valores ajustados e histograma dos resíduos da regressão linear dos dados de gambás das montanhas capturados na Inglaterra



Através do gráfico de dispersão dos resíduos versus valores preditos na Figura 8, notamos que os erros não aparentam ter nenhuma tendência, portanto, eles parecem ter uma variância constante, satisfazendo o critério de homoscedasticidade. Além disso, a distribuição dos resíduos aparenta ser unimodal, assemelhando-se a uma distribuição normal.

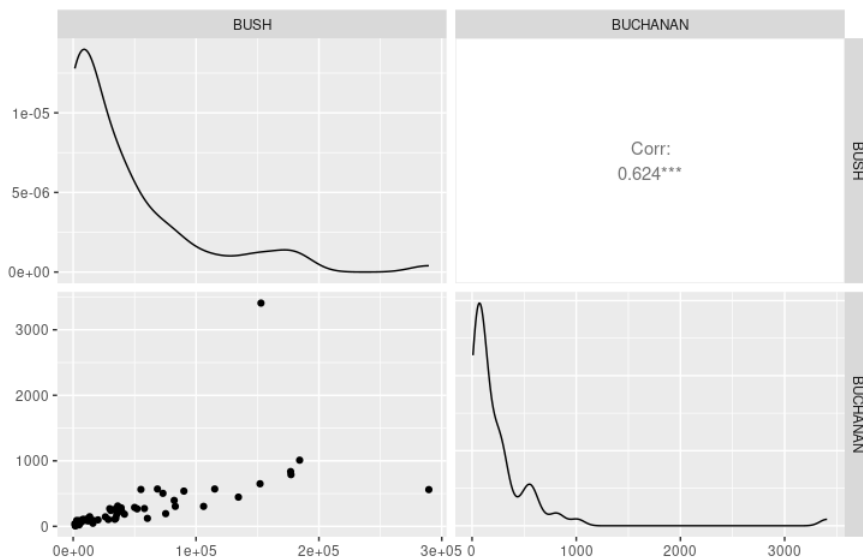
Vocês já realizaram análise de resíduos, mesmo sem ser solicitado... parabéns!

2 Eleição presidencial de 2000 no Estados Unidos

2.1 Relação entre os votos recebidos por Bush e por Buchanan

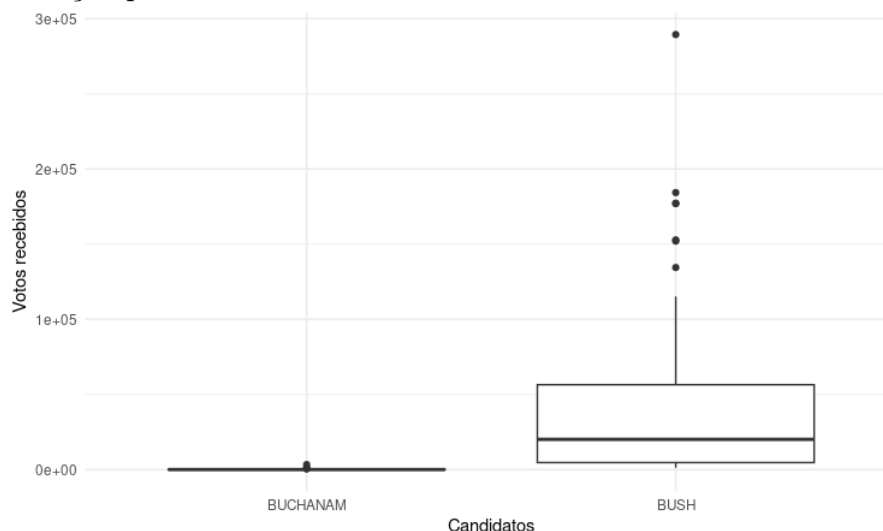
Olhando para o gráfico de dispersão na Figura 9, percebe-se que existe uma relação linear positiva entre os votos recebidos por Bush e por Buchanan. Entretanto, parece haver dois pontos atípicos que afetam negativamente essa visualização. Mas, o coeficiente de correlação de Pearson foi igual a 0,624, indicando uma correlação linear positiva moderada entre essas variáveis.

Figura 9: Correlação e gráficos de dispersão e densidade dos votos recebidos por condado, no estado da Flórida, por Bush e Buchanan na eleição presidencial dos Estados Unidos nos anos 2000



Os votos recebidos por Buchanan foram muito inferiores aos recebidos por Bush, enquanto o número máximo de votos recebidos em um condado de Buchanan foi 3.407, o número máximo recebidos por Bush foi 289.456.

Figura 10: *Boxplots* dos votos recebidos por condado, no estado da Flórida, por Bush e Buchanan na eleição presidencial dos Estados Unidos nos anos 2000



2.2 Ajuste um modelo de regressão linear

As estimativas obtidas foram $\hat{\beta}_0 = 45,29$ e $\hat{\beta}_1 = 0,0049$. O valor estimado para o β_1 indica que se existe uma relação linear, ela é positiva, pois o valor estimado foi maior que 0. O β_0 estimado nos informa que se a quantidade de votos recebidos por Bush for igual a 0, a quantidade de votos recebidos por Buchanan será de 45,29 (ou 45, se arredondarmos para o inteiro mais próximo).

Figura 11: Gráfico de dispersão e curva de regressão linear ajustada aos votos recebidos por condado, no estado da Flórida, por Bush e Buchanan na eleição presidencial dos Estados Unidos nos anos 2000

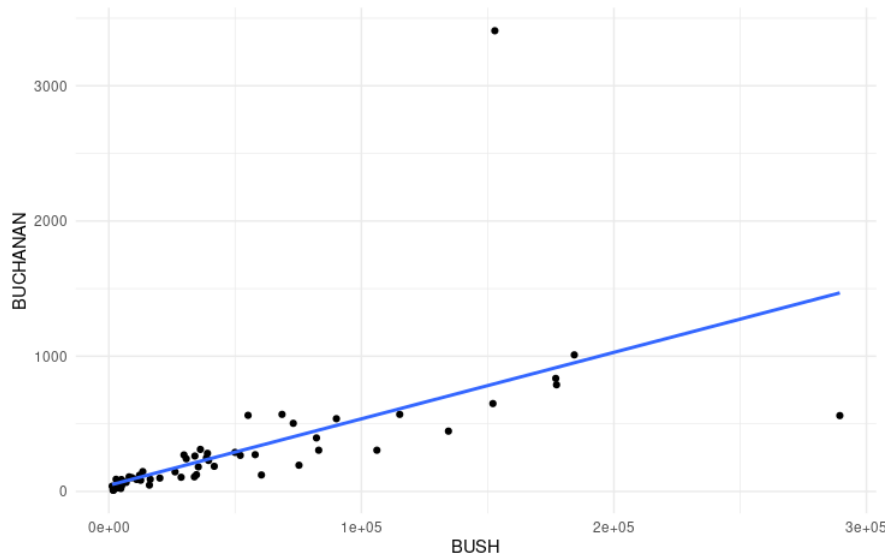
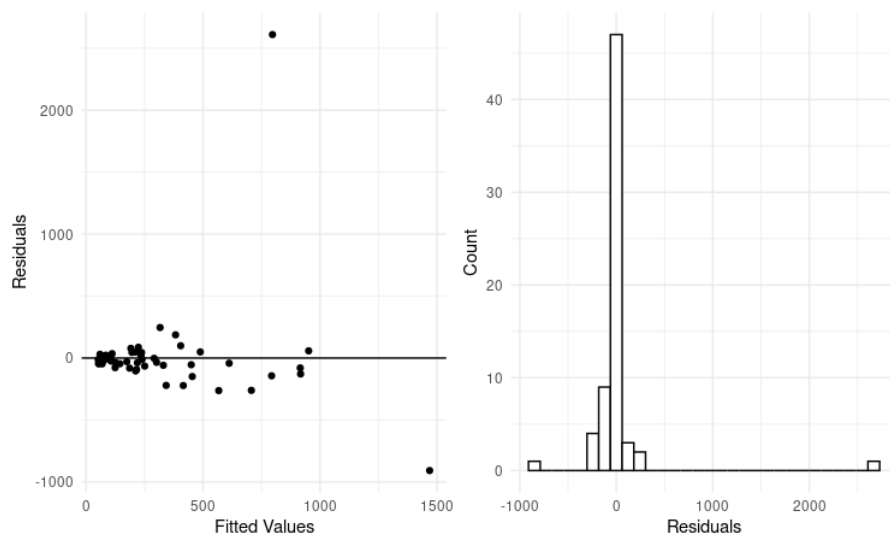


Figura 12: Gráfico dos resíduos versus valores ajustados e histograma dos resíduos da regressão linear dos votos recebidos por condado, no estado da Flórida, por Bush e Buchanan na eleição presidencial dos Estados Unidos nos anos 2000



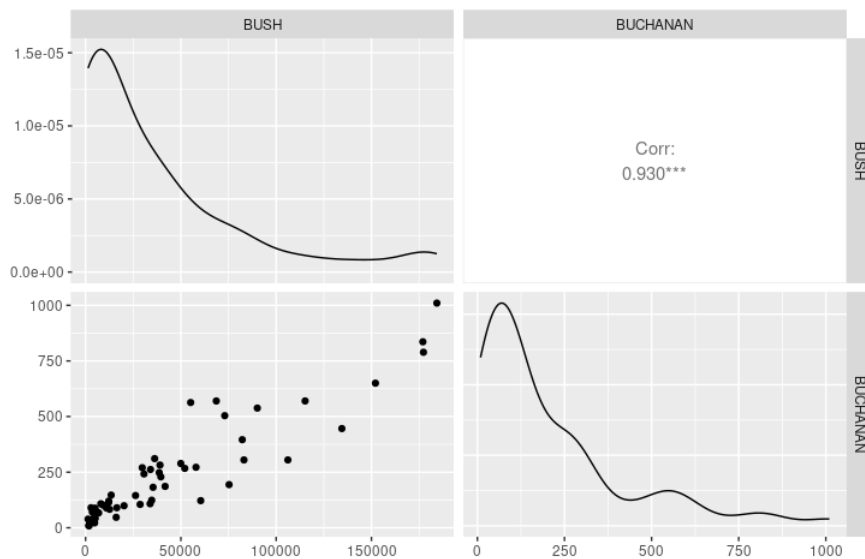
Observando o gráfico dos resíduos versus valores preditos na Figura 12, é possível notar que parece haver um aumento da variação nos resíduos à medida que os valores preditos são maiores, o que viola a suposição de homoscedasticidade do modelo de regressão linear.

Ademais, a distribuição dos resíduos não aparenta seguir uma distribuição normal, devido a presença de valores extremos.

2.3 Ajuste um modelo de regressão linear sem a presença de pontos atípicos

Como dito anteriormente, é possível perceber que há presença de valores atípicos nos dados e, portanto, esses valores foram retirados e as análises anteriores foram refeitas.

Figura 13: Correlação e gráficos de dispersão e densidade dos votos recebidos por condado, no estado da Flórida, por Bush e Buchanan na eleição presidencial dos Estados Unidos nos anos 2000



Agora com o novo gráfico de dispersão na Figura 13, é possível notar, de modo mais nítido, a relação linear entre os votos recebidos por Bush e Buchanan. O coeficiente de correlação de Pearson aumentou consideravelmente para 0,93, indicando uma forte correlação positiva linear entre essas variáveis.

As novas estimativas obtidas foram $\hat{\beta}_0 = 38,54$ e $\hat{\beta}_1 = 0,0044$. O valor estimado para o β_1 não foi tão alterado, continuando a nos indicar que se existe uma relação linear, ela é positiva, pois o valor estimado foi maior que 0. O novo β_0 estimado nos diz que se a quantidade de votos recebidos por Bush for igual a 0, a quantidade de votos recebidos por Buchanan será de 38,54 (ou 39, se arredondarmos para o inteiro mais próximo), o que é menor do que $\hat{\beta}_0$ anterior.

Observando o novo gráfico dos resíduos versus valores preditos na Figura 15, notamos que ainda assim parece que a variabilidade dos resíduos está aumentando para maiores valores preditos, violando a suposição de homoscedasticidade. Entretanto, a distribuição dos resíduos se assemelha mais a uma distribuição normal do que antes. Portanto, apesar do alto coeficiente de correlação de Pearson entre as variáveis, transformações precisariam ser feitas para que a suposição de homoscedasticidade fosse garantida.

Utilizando as estimativas obtida, é possível afirmar que se Bush receber 152.846 votos em determinado condado, Buchanan espera receber cerca de 712 votos nesse mesma área.

Sem o auxílio de funções, as estimativas de mínimos quadrados para o intercepto e o coeficiente angular do modelo de regressão linear simples foram as mesmas obtidas com a função *lm()*.

Parabéns
pela análise
de resíduos.

Figura 14: Gráfico de dispersão e curva de regressão linear ajustada aos votos recebidos por condado, no estado da Flórida, por Bush e Buchanan na eleição presidencial dos Estados Unidos nos anos 2000

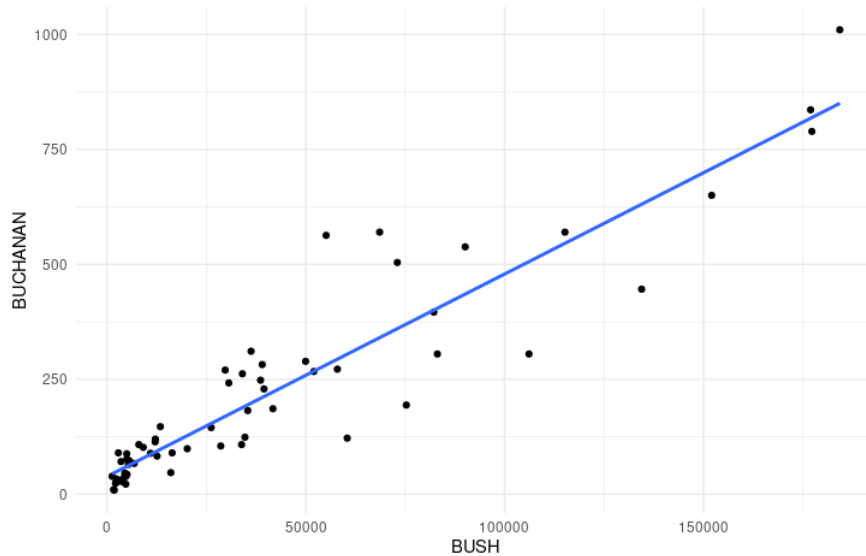
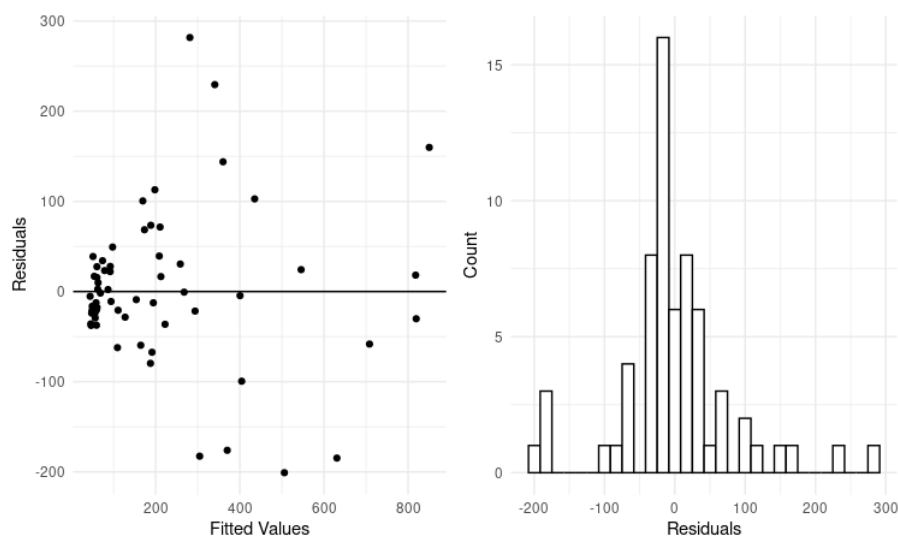


Figura 15: Gráfico dos resíduos versus valores ajustados e histograma dos resíduos da regressão linear dos votos recebidos por condado, no estado da Flórida, por Bush e Buchanan na eleição presidencial dos Estados Unidos nos anos 2000



Ainda sinto falta de um parágrafo final fechando o trabalho, indicando/discutindo os principais achados ou indicando propostas de continuidade da análise.