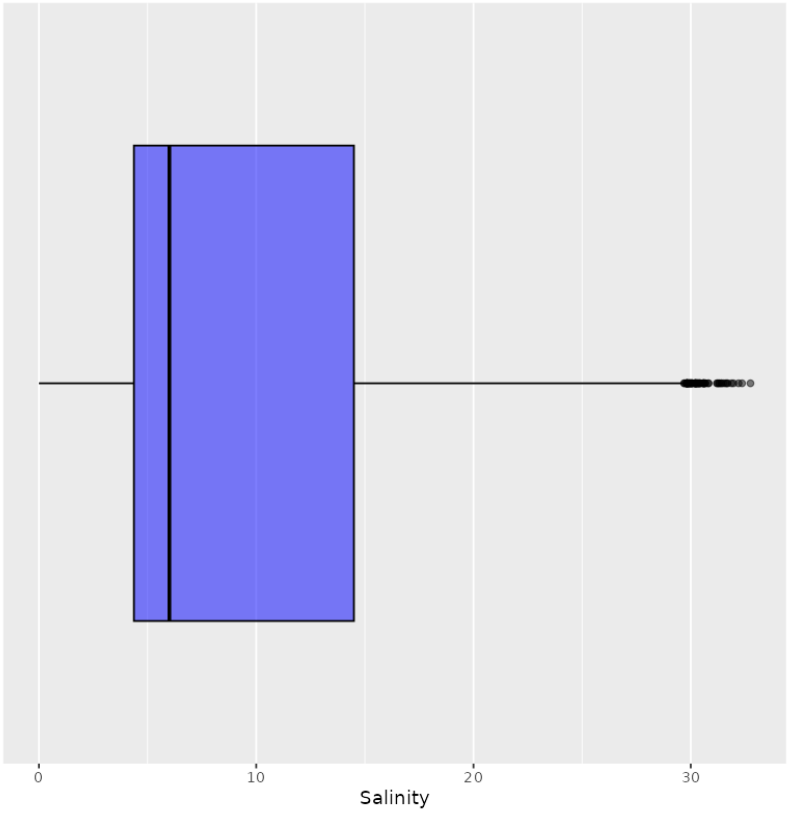
GEA1000 Tutorial 3

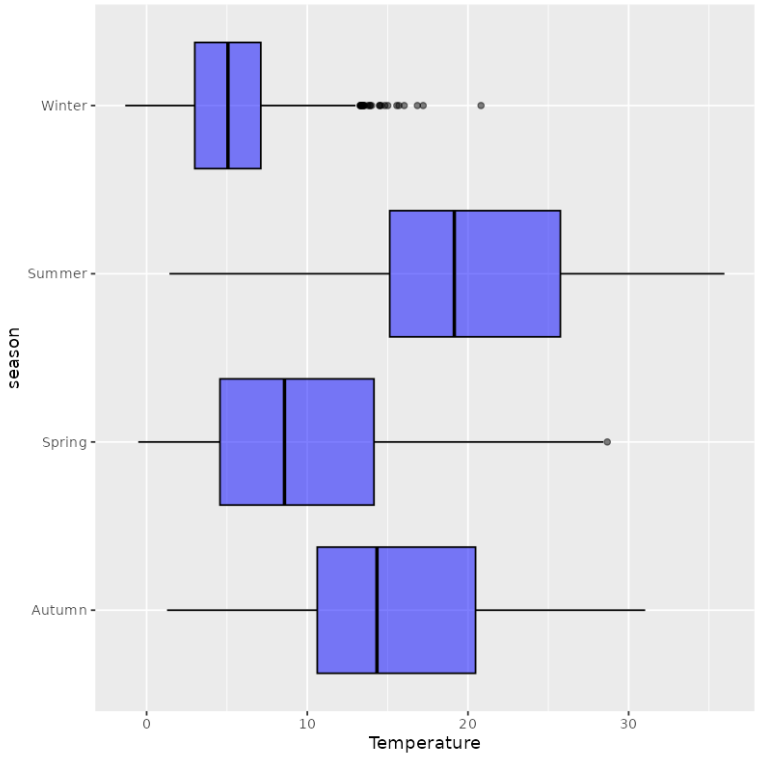
AY 24/25 Sem 2 — github/omgeta

Q1. (a.) There are 63 outliers.

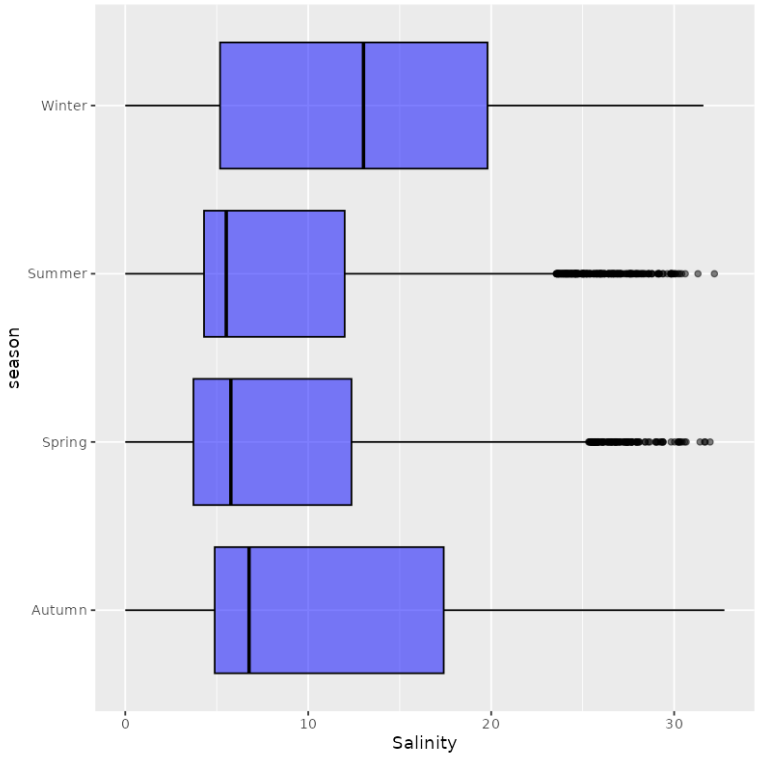
Statistic	Value
Mean	9.240
Median	6.005
Minimum	0
Maximum	32.748
SD	7.533
Q1	4.372
Q3	14.497
IQR	10.125



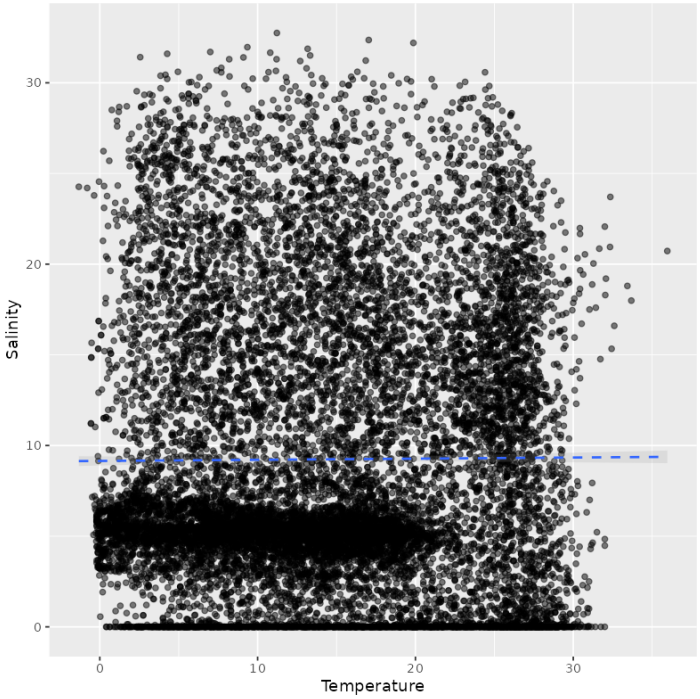
(b.) Summer has the largest IQR for temperature, Winter has the smallest IQR for temperature.



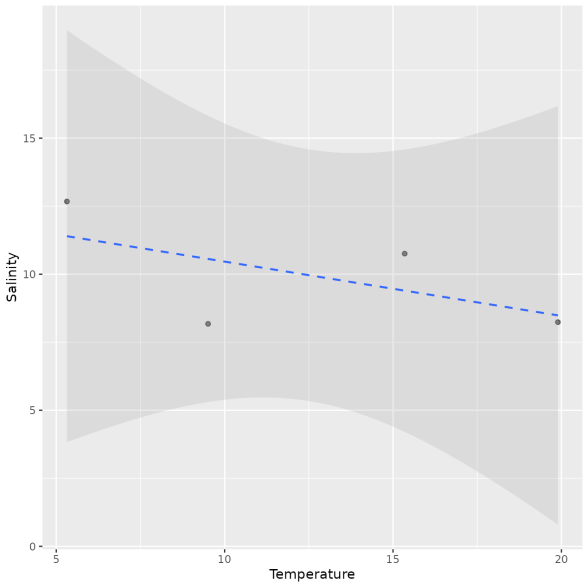
(c.) Winter has the largest IQR for salinity, Summer has the smallest IQR for salinity.



(d.) Correlation coefficient, $r = 0.01$. There is very small positive correlation.



(e.) Solution:

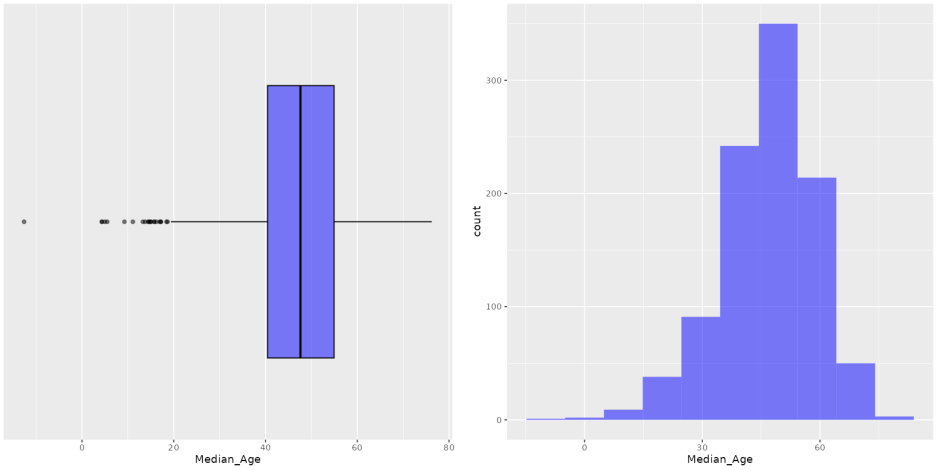


(f.) Ecological fallacy

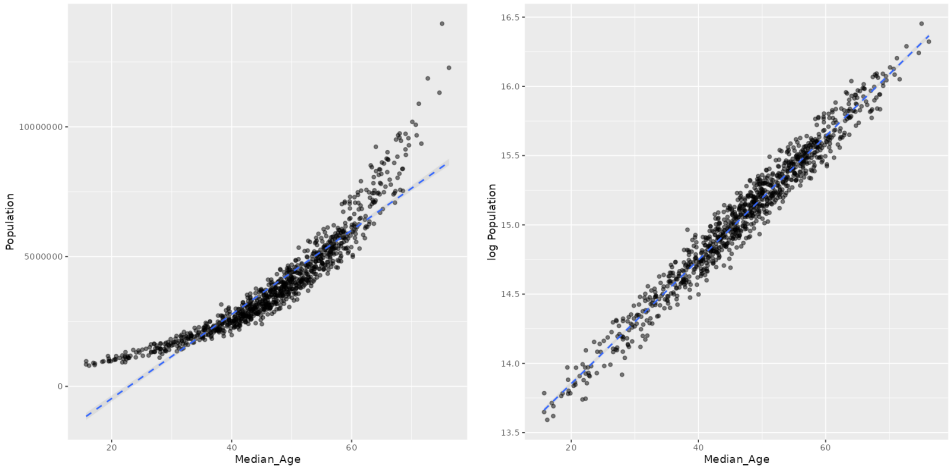
(g.) Salinity and temperature may have different relationships dependent on the season.

Q2. (a.) There are some outliers which have 0 or negative age.

Statistic	Value
Mean	46.705
Median	47.615
Minimum	-12.620
Maximum	76.230
SD	11.844
Q1	40.450
Q3	54.943
IQR	14.492



(b.) There is a position correlation between Population and Median_Age. Using the natural logarithm, we can normalize data which follows a natural exponential scaling to more accurate fit the correlation to Median_Age.



(c.) $\ln \text{Population} = 0.045 \times \text{Median\_Age} + 12.96$

(d.) Slope 0.045 gives the direction for greatest change between Median_Age and ln Population. Intercept 12.96 gives us the baseline for ln Population at Median_Age 0 which is not meaningful.

(e.) Predicted ln Population = 14.524. Therefore, predicted population is $e^{14.524} = 2030921$