Final Project Report
Big Data Analysis / BIO 539

AGUS SETIYAWAN (100542527)

The final project data was coming from observation on-board on pole-and-line vessel. The data set was collected from January to May 2015. The length of vessel is about 15 meters, depth is about 4 meters and made it by wood. This observation based on two fishing ground including Maluku Waters and Sulawesi Waters in Indonesia.

Table 1. Data Summary from January to May 2015

No	Month	Fishing Ground	Live Bait	Total Bait (kg)	Total fishing	Total Catch (kg)	Average total bait (kg)	Average Total Catch (kg)
1	January	Maluku Waters	Anchovy	354.6	9	3990	39.40	443.33
2	February	Maluku Waters	Red Anchovy	2402	23	14055	104.43	611.09
			Round scad and Red Anchovy	1371	7	9784	195.86	1397.71
3	March	Maluku Waters	Red Anchovy	34.2	3	366	11.40	122.00
			Round scad and Red Anchovy	129.6	4	2473	32.40	618.25
			Round scad	470	14	6523	33.57	465.93
4	April	Sulawesi Waters	Anchovy	298.8	9	4700	33.20	522.22
5	May	Maluku Waters	Anchovy	857	8	12307	107.13	1538.38
	Total			5917.20	77	54198	557.39	5718.91

Note: Data was collected by observer program.

Based on that summary data set, the objective for this research is about the influence factors to CPUE for pole-and-line fishery in Indonesian waters. As mentioned before, that pole-and-line fishery is depending on the availability of live-bait, could say that no live-bait equal no fishing. The fishing technique of pole-and-line is due to the fishing activities, the live-bait is released to attract / fake bait makes an easy to fishing.

The table 2 below is explain that GLM (Generalized Linear Model) analysis on this research. Before the analysis, the variable is necessary to understand which as factor or as numerical.

Table 2. GLM (*Generalized Linear Model*) Data Set Summary.

Factors	Level	Category	Type
Fishing ground	1	Maluku Waters	Categorical
	2	Sulawesi Waters	
Fishing Time	1	05.00 AM - 09.59 AM	Categorical
	2	10.00 AM - 13.59 PM	
	3	14.00 PM - 18.00 PM	
Month	1	January	Categorical
	2	February	
	3	March	
	4	April	
	5	May	
Live bait	1	Anchovy	Categorical
	2	Round scad	
	3	Red Anchovy	
	4	Red Anchovy and Round scad	
Fishing Duration	-	-	Continuous
Sea Surface Temperature	-	-	Continuous
Total fishers	-	-	Continuous
CPUE	-	-	Continuous
Total Catch	-	-	Continuous

Based on the summary data for GLM analysis, the variables as factor are Fishing ground, Fishing time, Month and Live bait. The variables as a numeric are Fishing duration, Sea Surface Temperature, Total fishers and CPUE.
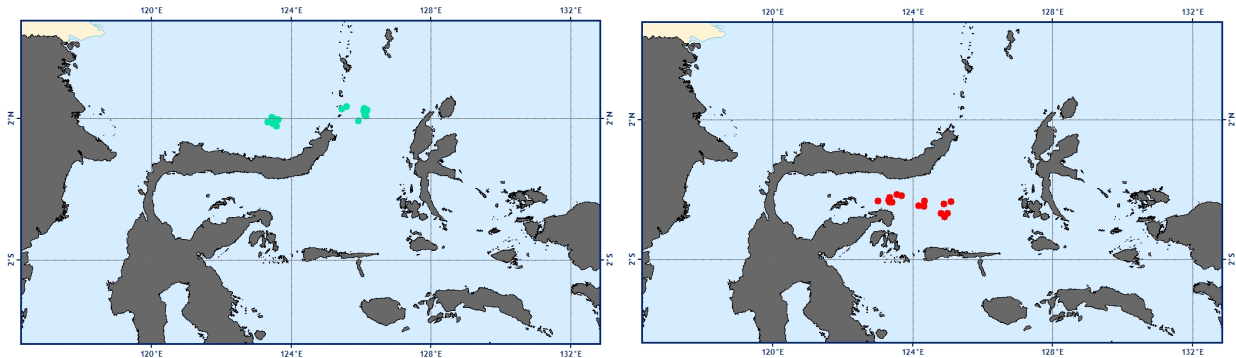
The equation for the GLM (*Generalized Linear Model*) analysis:

CPUE (Catch per Unit Effort) $= \beta_{1j}Fishing\_Duration_{ij} + \beta_{2j}Live\_Bait_{ij} + \beta_{3j}Fishing\_Ground_{ij} + \beta_{4j}Fishing\_Time_{ij} + \beta_{4j}Fishing\_Time_{ij} + \beta_{5j}Month_{ij} + poly~(SST),$ *family = gaussian.*
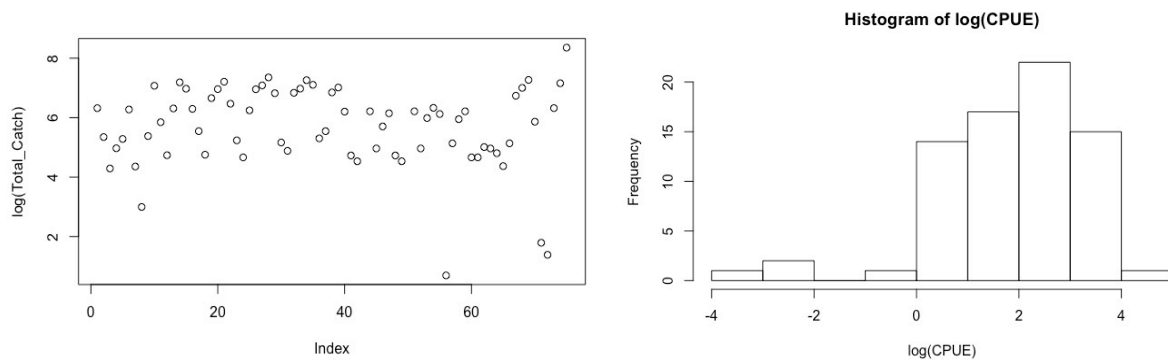
RESULT

1. **Fishing Ground**

That figures are displayed the fishing ground from 2 fishing area, left side is called Sulawesi Waters and right side is called Maluku Waters.

2. **Total Catch Distribution and CPUE**

That figures are displayed the distribution of total catch and CPUE from January – May 2015. Based on the log total catch we could inform that the total catch is a normal distribution over 5 Month. Interestingly, based on the log CPUE histogram inform that any high CPUE on the trip from January to May 2015.

### 3. Boxplot by Total catch and CPUE.
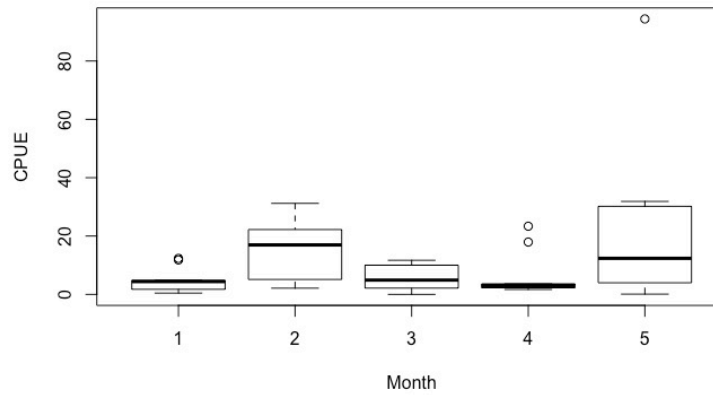
Based on the boxplot figure, we could inform that any differences CPUE on Month May, the hypothesis we could say that on May the vessel has caught a fish with less for effort, if it compares with the other Month.

4. GLM (*Generalized linear Model*) Analysis.

Coefficients:	Estimate	Std. Error	t value	Pr(>|t|)	Signif.code
(Intercept)	-5.69429	203.65634	-0.028	0.97778	
poly(SST)	-3.33135	5.03650	-0.661	0.51074	
as.factor(Month)2	-3.98370	3.95317	-1.008	0.31744	
as.factor(Month)3	-1.08726	1.99593	-0.545	0.58785	
as.factor(Month)4	3.85791	4.01649	0.961	0.34047	
as.factor(Month)5	8.50089	1.93101	4.402	4.23e-05	***
as.factor(Time)2	2.65236	1.57643	1.683	0.09742	.
as.factor(Time)3	3.83072	1.24860	3.068	0.00318	**
Long	-0.00769	1.61855	-0.005	0.99622	
Lat	-0.13063	0.14003	-0.933	0.35444	
Fishing_Duration	0.19333	0.04051	4.772	1.12e-05	***
Bait_Amount	0.11265	0.01115	10.107	8.26e-15	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Based on the analysis, we could inform that the factors may influence to the CPUE are Month 5 (May), Fishing time between 2 and 3 (10.00 am – 16.00 pm), Fishing duration and Bait amount.
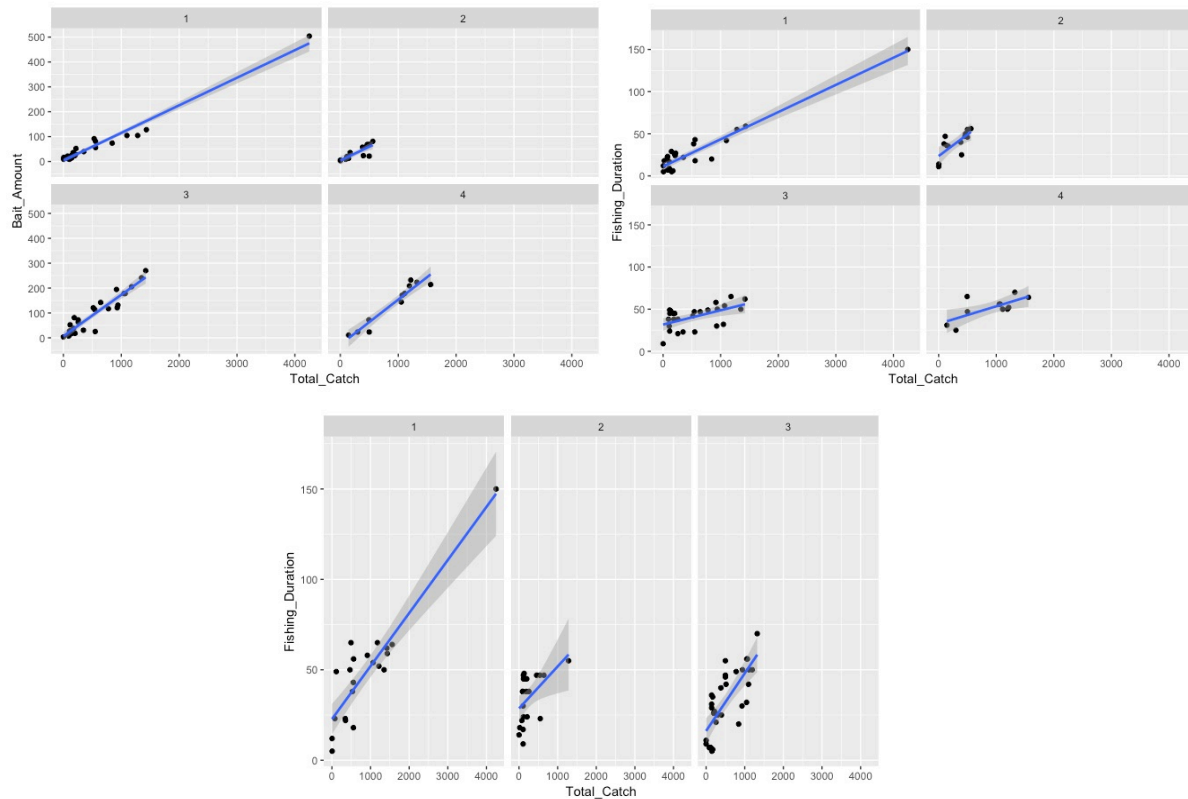

Those graphs are plot by ggplot and it informed that based on the 4 type of a live-bait that the data fitted to live-bait 4, it means that the live-bait mixed among Round scad and Red anchovy.

5. The Best Live-Bait to use for Pole-and-Line Fishery

Coefficients:	Estimate	Std. Error	t value	Pr(>|t|)	Signif.code
(Intercept)	509.04	121.09	4.204	7.54e-05	***
as.factor (Live_Bait)2	-217.96	207.03	-1.053	0.2960	
as.factor (Live_Bait)3	28.69	169.60	0.169	0.8661	
as.factor(Live_Bait)4	396.69	219.07	1.811	0.0744	.

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Based on the result, we could inform that the best live-bait to use for pole-and-line fishery is mixed between Round scad and Red anchovy with significantly P > 0.0744.

Normal Q-Q

lm(Total_Catch ~ as.factor(Live_Bait))
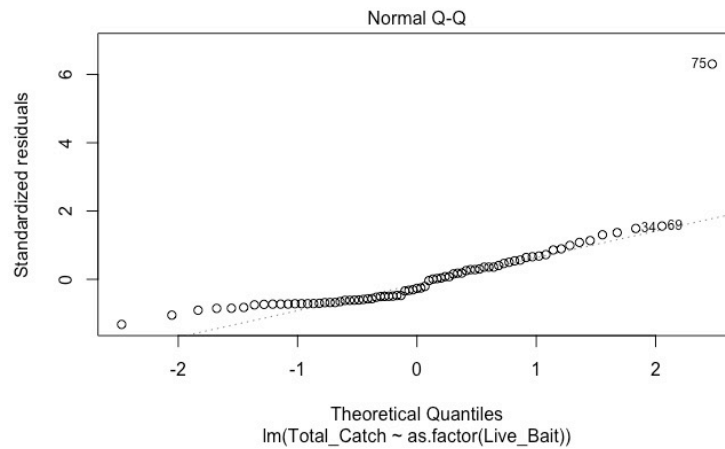
That graph meaning is how the data has fitted, based on the analysis by linear model for Total catch to live-bait.

## 6. The Linear Regression of Total catch
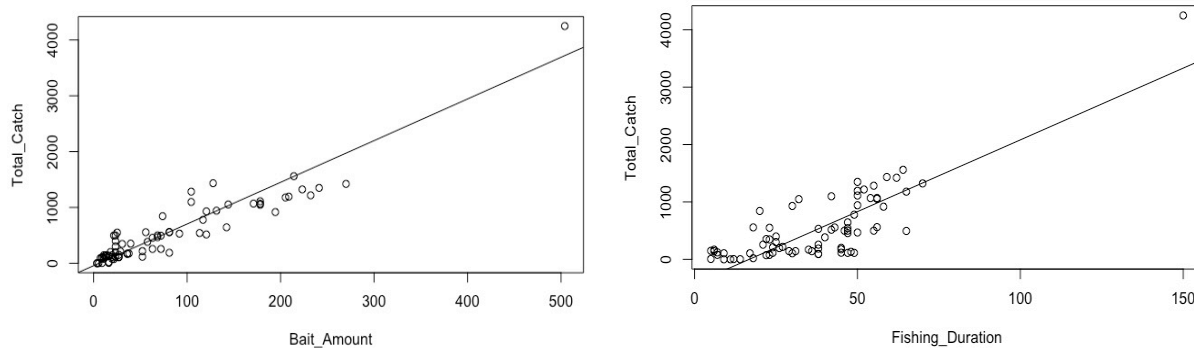
The correlation between bait amount and total catch is 0.9356093 (93%), it means that the relationship is high fitted to the data. Similarly, the result of the correlation between fishing duration and total catch is 0.81423 (81%), it means that the relationship is high fitted to the data.

7. Plotting by ggplot (Total catch vs Bait amount)
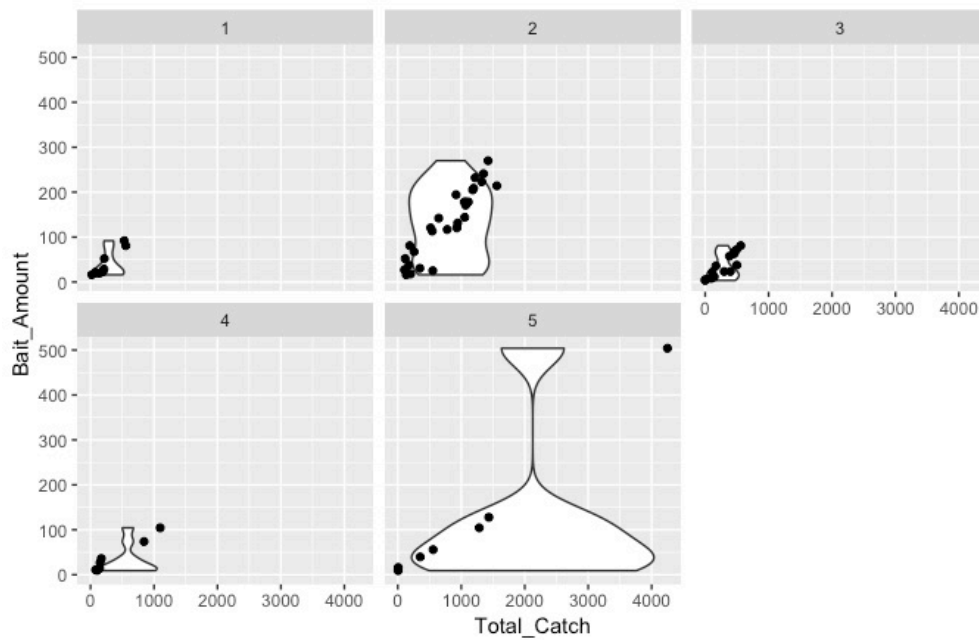
      That graph is displayed the relationship between total catch and bait amount, it based Month from January to May 2015, we could inform that any big fishes has caught by pole-and-line on May 2015.
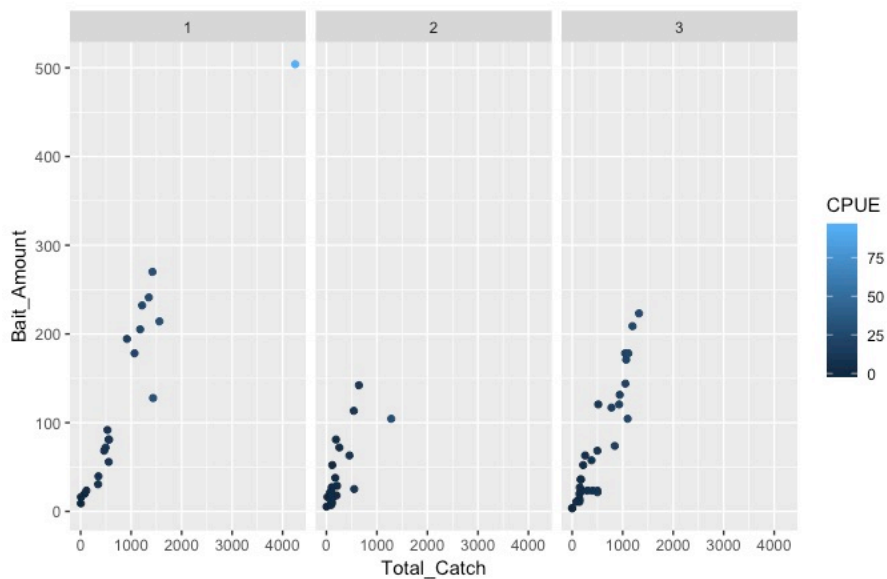
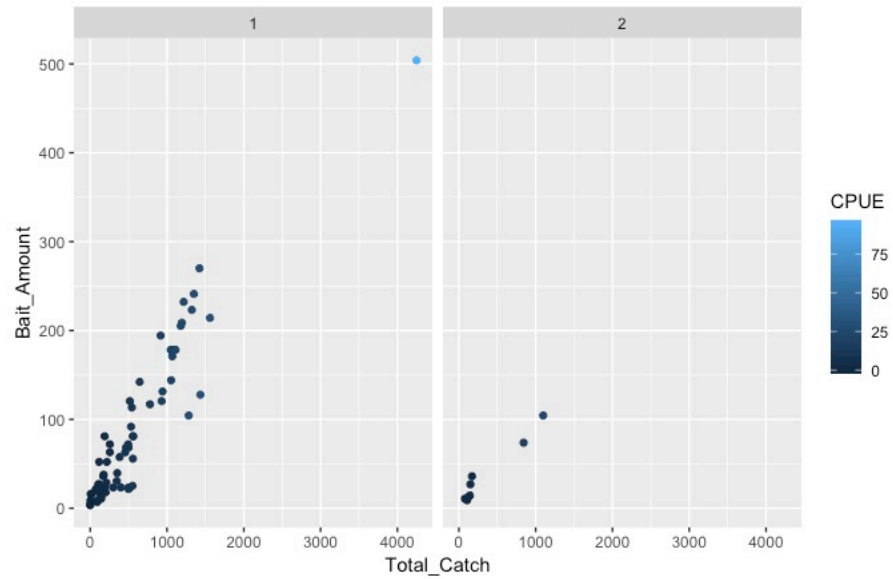8. Plotting by ggplot develop the function a (x = Total Catch, y = Bait Amount)
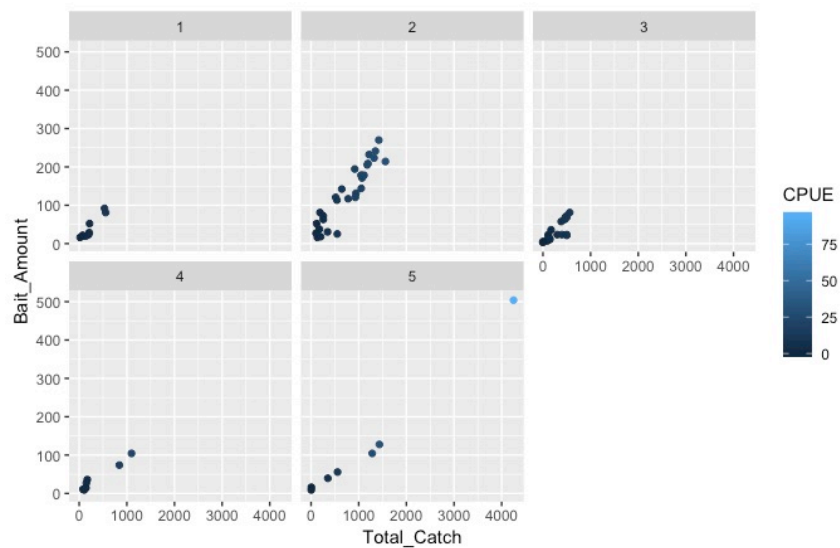
That graph is displayed the relationship between the total catch and bait amount, it based on the Month from January to May 2015.
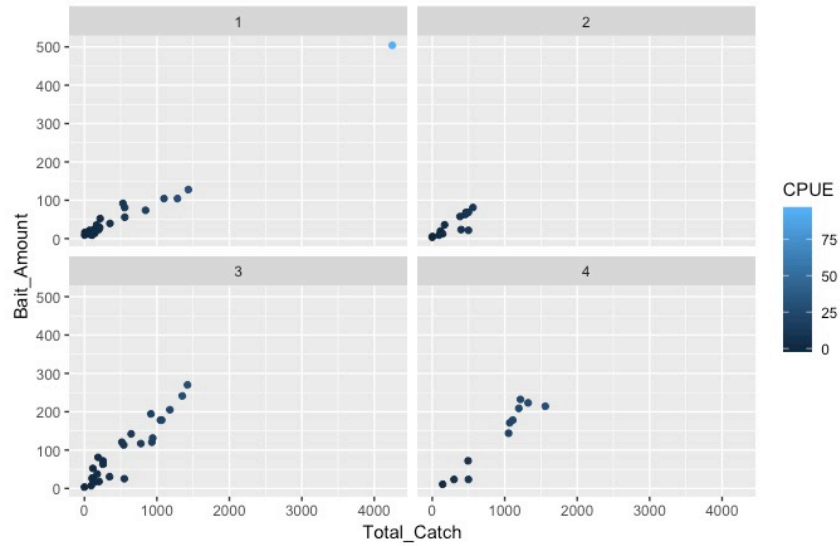
That graph is displayed the relationship between Total catch and Bait amount, it based on the fishing time over 5 Month from January to May 2015.
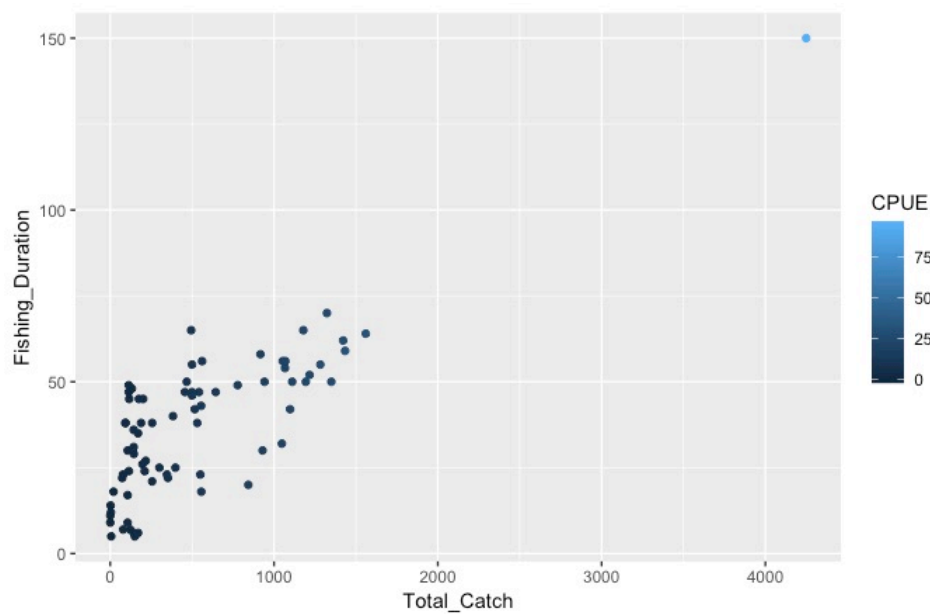

That graph is displayed that relationship between total catch and Bait amount, it based on the fishing time over 5 Month from January to May 2015.
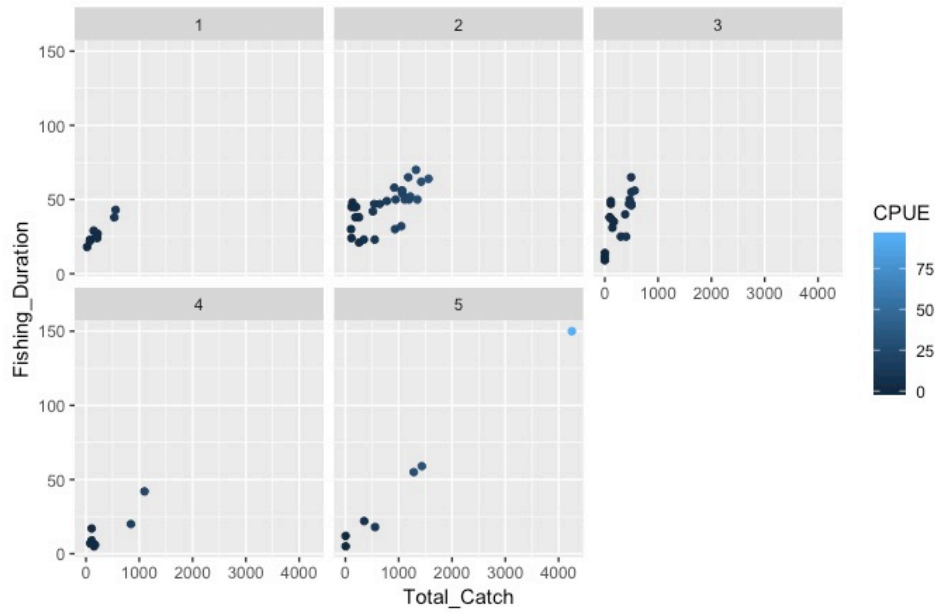
That graph is displayed the relationship between Total catch and Bait amount, it based on The Fishing ground (Maluku Waters and Sulawesi Waters) over 5 Month from January to May 2015.
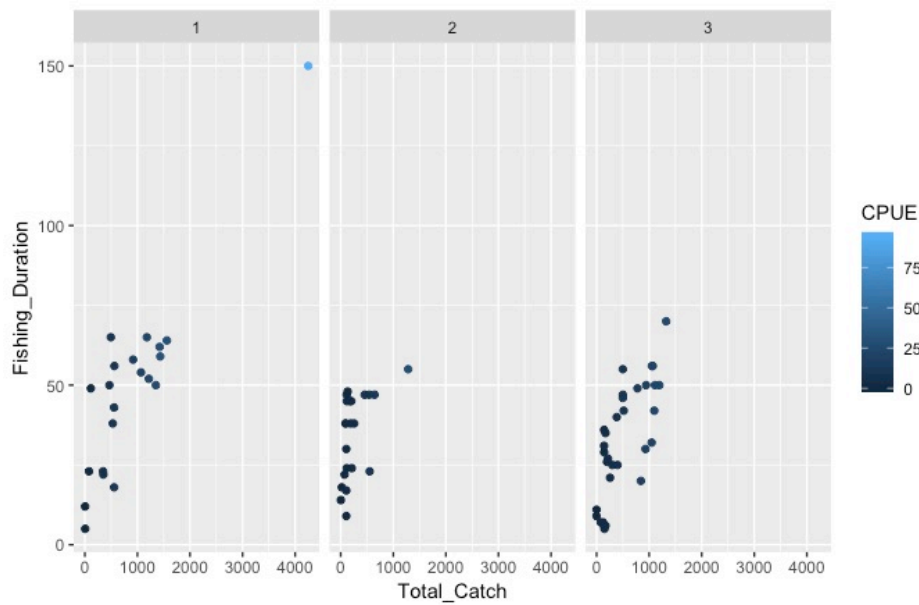
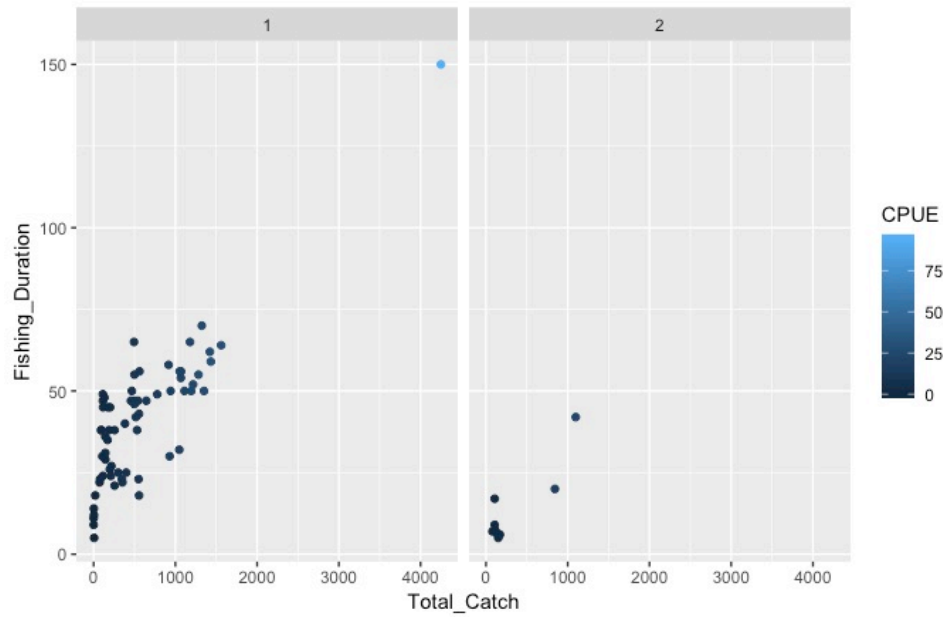9. Plotting by ggplot (Total catch vs Fishing Duration)

That graph is displayed the relationship between Total catch and Fishing duration over 5 Month from January to May 2015.

That graph is displayed the relationship between Total catch and Fishing duration, it based on the 5 Month from January to May 2015.

That graph is displayed the relationship between Total catch and fishing duration, it based on the fishing time over 5 Month from January to May 2015.

That graph is displayed the relationship between Total catch and Fishing duration, it based on the Fishing ground (Maluku waters and Sulawesi waters) over 5 Month from January to May 2015.