

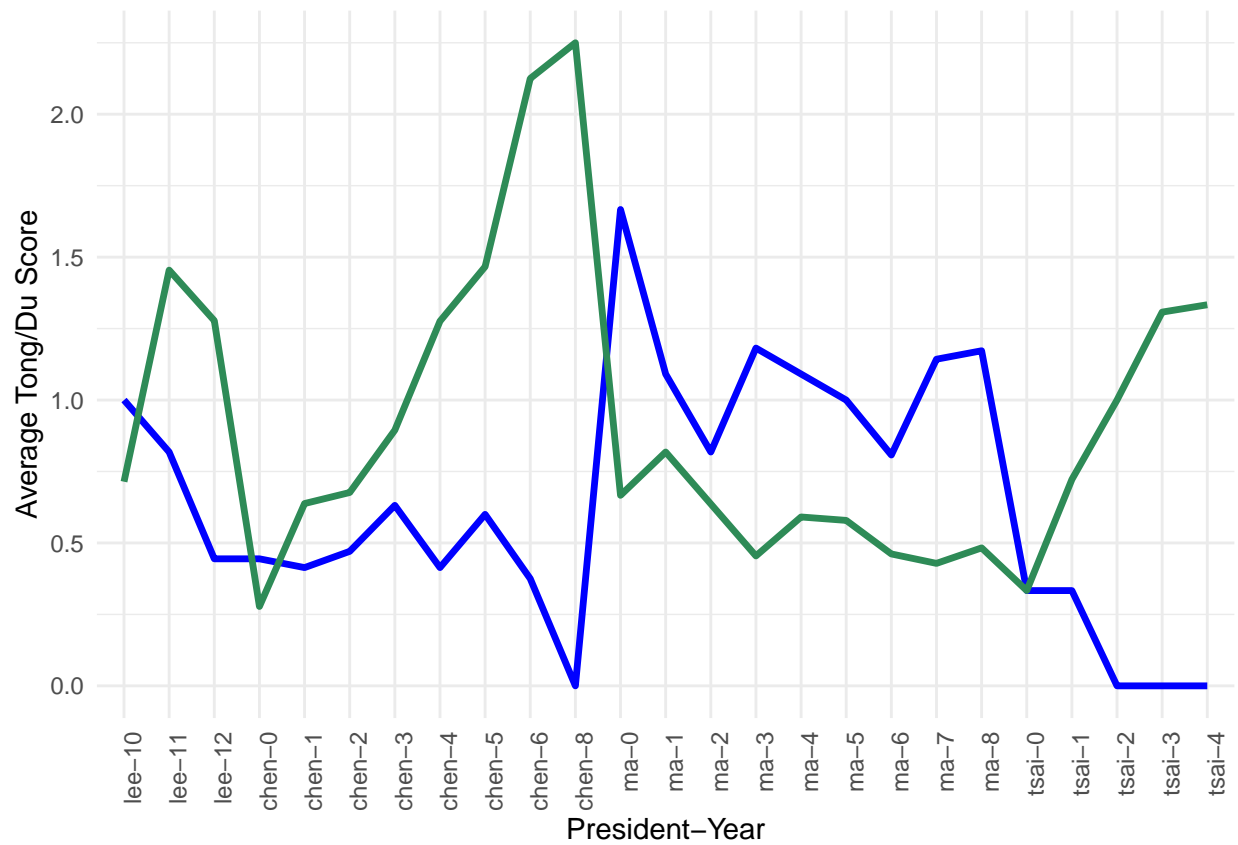
Aggregating TongDu Tables

Maggie X. Wang

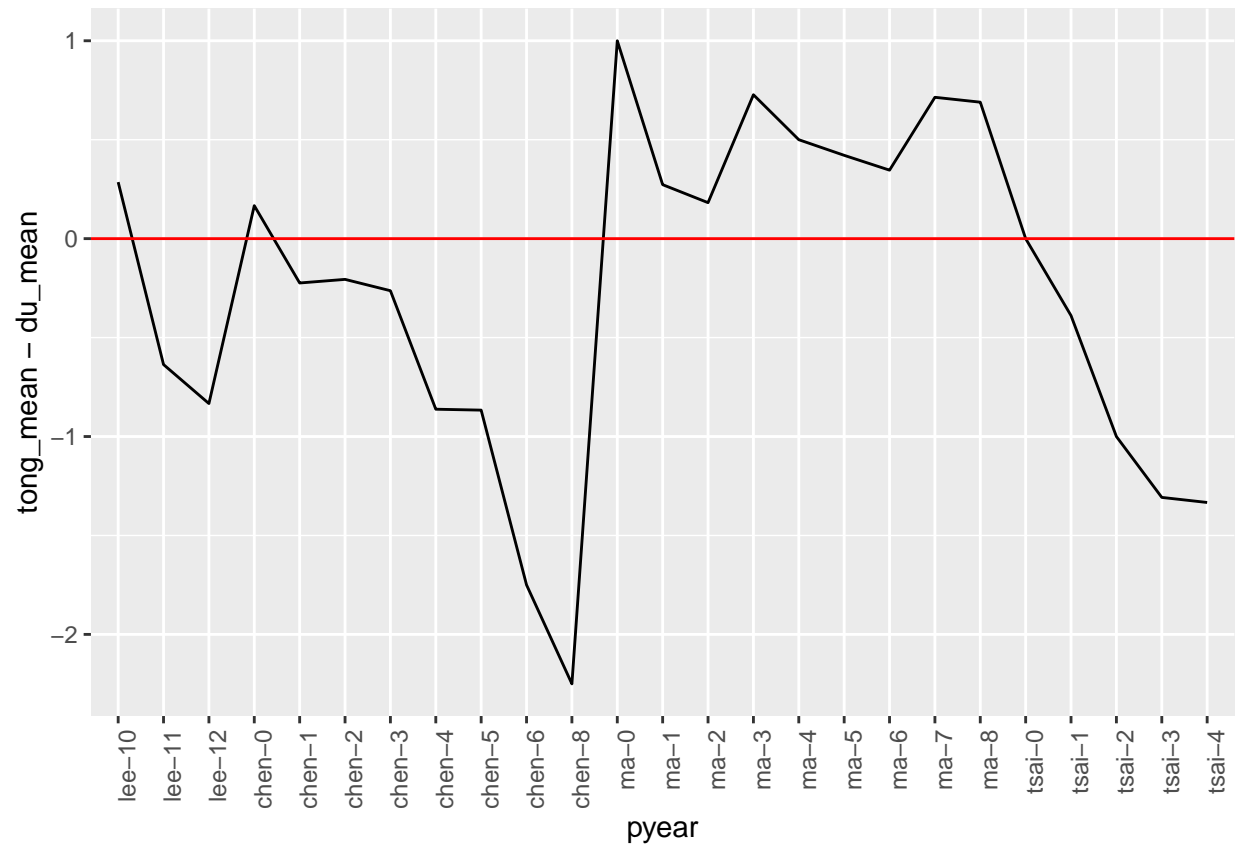
2/21/2020

```
#aggregation1
tongdu_year <- sqldf('SELECT *
FROM tongdu_event
LEFT JOIN president_year ON ((date BETWEEN start and end) AND (tongdu_event.president=py_president))') %>%
  filter(is.na(exclude)) %>%
  filter(source=="mac") %>%
  mutate(pyear = forcats::fct_explicit_na(pyear)) %>%
  group_by(pyear) %>%
  summarise(n = n(),
            tong_mean = mean(tong, na.rm=TRUE),
            du_mean = mean(du, na.rm = TRUE),
            tong_total = sum(tong, na.rm = TRUE),
            du_total = sum(du, na.rm = TRUE))

ggplot(tongdu_year, aes(x=pyear)) +
  #geom_bar(stat = "identity", aes(y=n/15)) +
  geom_line(aes(y=tong_mean, group = 1), color = "blue", size = 1.2) +
  geom_line(aes(y=du_mean, group = 1), color = "seagreen", size = 1.2) +
  xlab("President-Year") +
  ylab("Average Tong/Du Score") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))
```

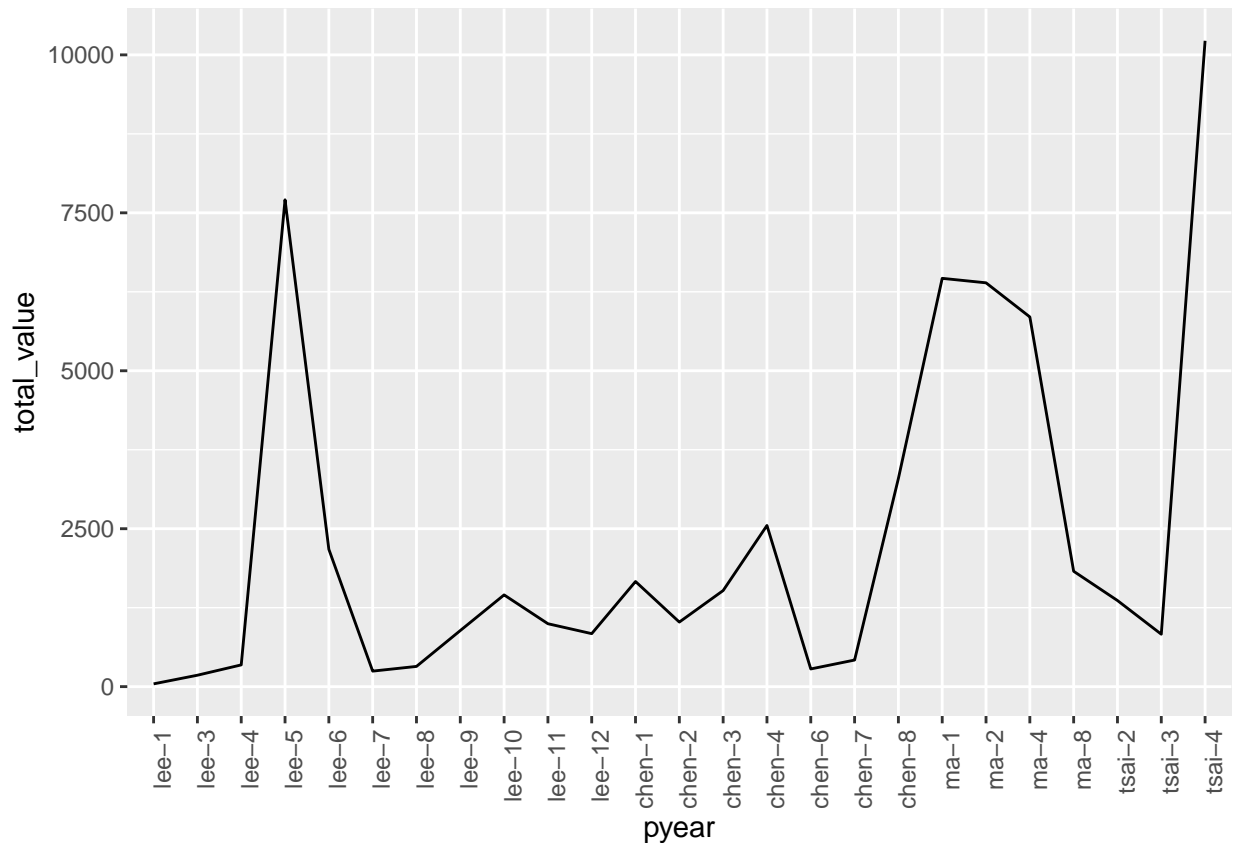


```
ggplot(tongdu_year, aes(x=pyear, y=tong_mean-du_mean)) +
  geom_line(aes(group = 1)) +
  geom_hline(yintercept=0, color = "red") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))
```



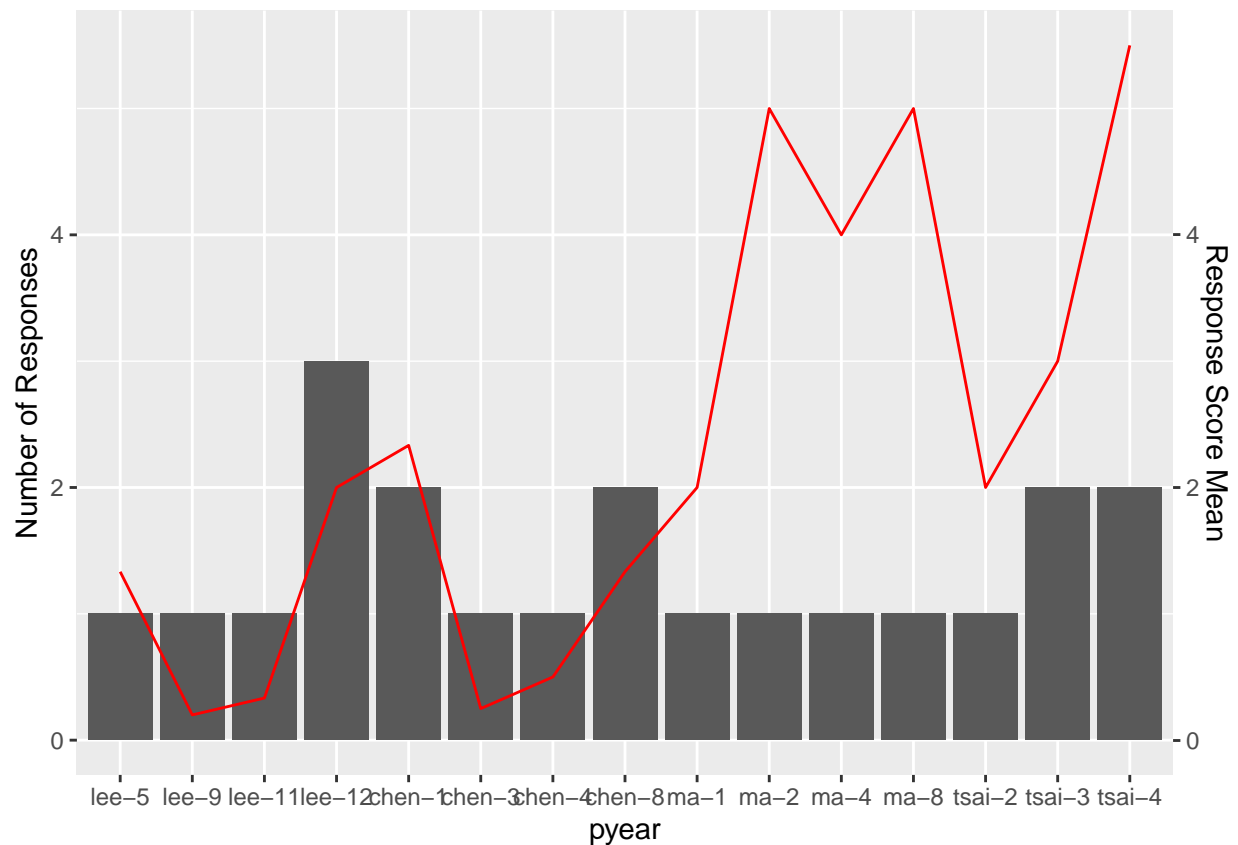
```
#aggregation2: arms sales with pyear
transaction_year <- sqldf('SELECT *
FROM transactions
LEFT JOIN president_year ON date BETWEEN start and end') %>%
  group_by(pyear) %>%
  summarise(n = n_distinct(deal_id),
            total_value = sum(value_million, na.rm = TRUE),
            mean_value = mean(value_million, na.rm = TRUE))
```

```
ggplot(transaction_year, aes(x=pyear, y=total_value)) + geom_line(aes(group = 1)) +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))
```

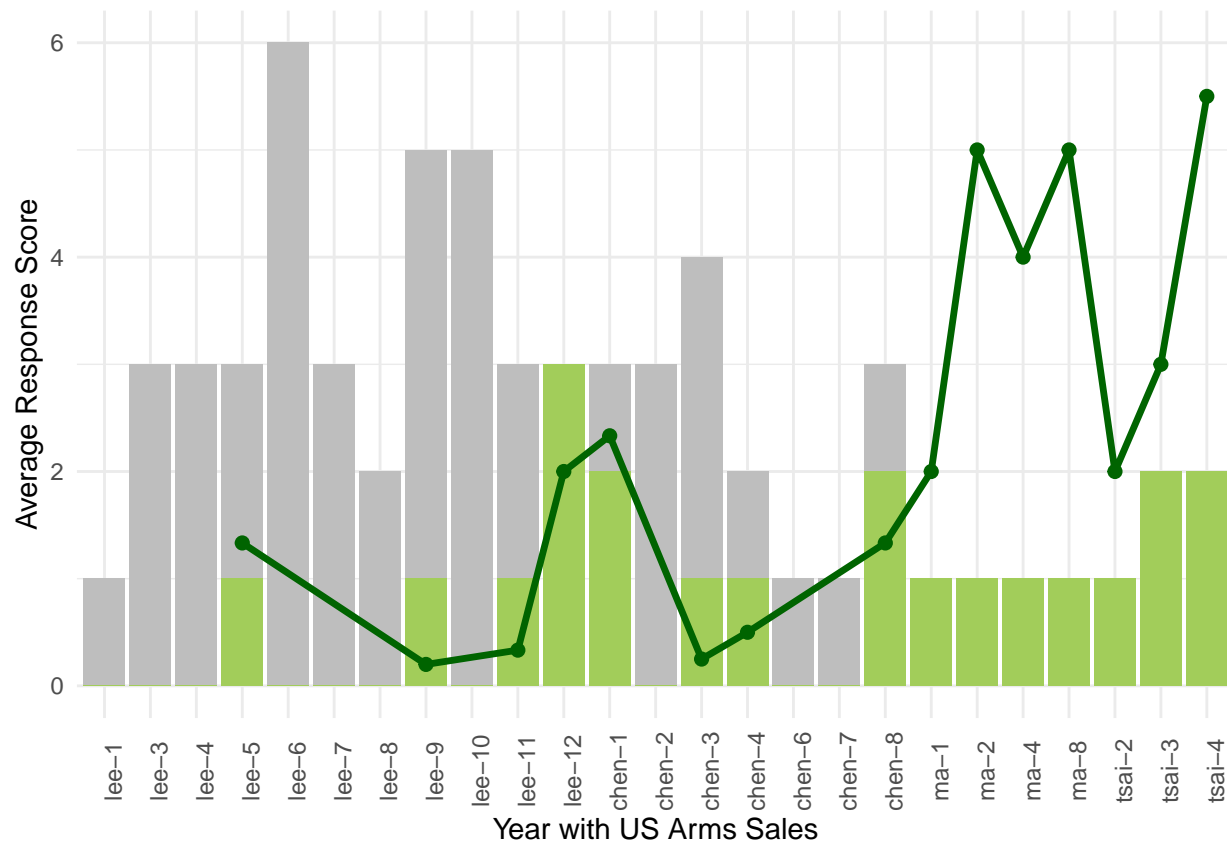


```
#aggregation3: transaction-reaction with pyear
reaction <- sqldf('SELECT *
FROM transaction_reaction
LEFT JOIN president_year ON date BETWEEN start and end') %>%
  group_by(pyear) %>%
  summarise(total_us = sum(target_us, na.rm = TRUE),
            total_tw = sum(target_tw, na.rm = TRUE),
            response_mean = mean(response, na.rm = TRUE),
            response_total = sum(response, na.rm = TRUE),
            n_response = length(which(response!=0)))
```

```
reaction %>%
  filter(response_total!=0) %>%
  ggplot() +
  geom_bar(stat = "identity", aes(x=pyear, y = n_response)) +
  geom_line(aes(x=pyear, y = response_mean, group = 1), color = "red") +
  scale_y_continuous(name = "Number of Responses",
                     sec.axis = sec_axis(~., name = "Response Score Mean"))
```



```
transaction_year %>%
  ggplot() +
  geom_bar(stat = "identity", aes(x=pyear, y=n), fill = "grey") +
  geom_bar(data = reaction, aes(x=pyear, y = n_response), stat = "identity", fill = "darkolivegreen3") +
  geom_line(data = filter(reaction, n_response!=0), aes(x=pyear, y = response_mean, group = 1), color = "red") +
  geom_point(data = filter(reaction, n_response!=0), aes(x=pyear, y = response_mean, group = 1), color = "red") +
  xlab("Year with US Arms Sales") +
  ylab("Average Response Score") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 90))
```



```
#grand aggregation
meta <- tongdu_year %>%
  left_join(reaction, by = "pyear") %>%
  left_join(select(transaction_year, -n), by = "pyear") %>%
  mutate(n_event = n,
         tongdu_diff = tong_mean-du_mean,
         tongdu_diff_ttl = tong_total-du_total) %>%
  select(-n)

#meta_response <- meta %>% filter(!is.na(response_mean)) %>%
# mutate(response_yn = ifelse(response_total==0, 0, 1))

meta_response <- meta %>% filter(n_response!=0)

m_yn <- lm(n_response ~ tong_mean + du_mean + n_event + total_value, data = meta_response)
summary(m_yn)

##
## Call:
## lm(formula = n_response ~ tong_mean + du_mean + n_event + total_value,
##     data = meta_response)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.76373 -0.29019 -0.00809  0.13906  1.44514
##
## Coefficients:
```

```
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  9.005e-01  1.221e+00   0.737   0.482
## tong_mean   -5.599e-01  5.629e-01  -0.995   0.349
## du_mean      4.884e-01  6.239e-01   0.783   0.456
## n_event      1.466e-02  1.795e-02   0.816   0.438
## total_value  1.835e-05  7.429e-05   0.247   0.811
##
## Residual standard error: 0.6503 on 8 degrees of freedom
## Multiple R-squared:  0.3533, Adjusted R-squared:  0.02992
## F-statistic: 1.093 on 4 and 8 DF,  p-value: 0.4221

m_tongdu <- lm(response_total ~ tong_mean + du_mean + total_value, data = meta)
summary(m_tongdu)
```

```
##
## Call:
## lm(formula = response_total ~ tong_mean + du_mean + total_value,
##     data = meta)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.2826 -1.6763 -0.0939  1.5663  3.7322
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  5.6281660  2.6943906   2.089   0.0587 .
## tong_mean    -3.8969324  1.9401702  -2.009   0.0676 .
## du_mean      -1.6725417  1.5463917  -1.082   0.3007
## total_value   0.0006063  0.0002332   2.600   0.0232 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.519 on 12 degrees of freedom
## (9 observations deleted due to missingness)
## Multiple R-squared:  0.4782, Adjusted R-squared:  0.3477
## F-statistic: 3.665 on 3 and 12 DF,  p-value: 0.04396
```

```
#trying out some regressions
m_tong <- lm(response_total ~ tong_mean + n_event + total_value, data = meta)
summary(m_tong)
```

```
##
## Call:
## lm(formula = response_total ~ tong_mean + n_event + total_value,
##     data = meta)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.3621 -1.3649 -0.2349  1.3765  3.8050
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.6602403  1.5069433   1.102  0.29219
## tong_mean    -3.0093988  1.5262055  -1.972  0.07213 .
## n_event       0.0693607  0.0474751   1.461  0.16970
```

```
## total_value 0.0007435 0.0002350 3.164 0.00816 **
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.432 on 12 degrees of freedom
## (9 observations deleted due to missingness)
## Multiple R-squared: 0.5138, Adjusted R-squared: 0.3922
## F-statistic: 4.227 on 3 and 12 DF, p-value: 0.02956

m_du <- lm(response_total ~ du_mean + n_event + total_value, data = meta)
summary(m_du)
```

```
##
## Call:
## lm(formula = response_total ~ du_mean + n_event + total_value,
##     data = meta)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.3696 -1.4989 -0.4915  1.8803  4.3071
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.9773905  3.1353061  -0.631  0.5401
## du_mean      1.4516864  1.6350901   0.888  0.3921
## n_event      0.0868265  0.0638138   1.361  0.1986
## total_value  0.0007749  0.0002777   2.790  0.0163 *
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.71 on 12 degrees of freedom
## (9 observations deleted due to missingness)
## Multiple R-squared: 0.3959, Adjusted R-squared: 0.2449
## F-statistic: 2.622 on 3 and 12 DF, p-value: 0.09869
```

```
m_diff <- lm(response_total ~ tongdu_diff + total_value, data = meta)
summary(m_diff)
```

```
##
## Call:
## lm(formula = response_total ~ tongdu_diff + total_value, data = meta)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.1048 -1.9333 -0.8441  2.5493  4.5856
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.1965858  1.1362476   1.053  0.3115
## tongdu_diff -0.7189143  0.8365996  -0.859  0.4057
## total_value  0.0006350  0.0002515   2.525  0.0254 *
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.723 on 13 degrees of freedom
```



```
## (9 observations deleted due to missingness)
## Multiple R-squared: 0.3394, Adjusted R-squared: 0.2378
## F-statistic: 3.339 on 2 and 13 DF, p-value: 0.06755
```