Summary of Trade and Tariff Data, Quantity Gap

September 05, 2017

This analysis is for the UN Comtrade trade data that matches with tariff rates from WITS for the HS 2012 classification, over years 2012-2016.

Notes:

- *This uses tariff data from WITS in ad valorem equivalent format. I donwloaded the AVE tariff data from the bulk download option at this page: http://wits.worldbank.org/WITS/WITS/AdvanceQuery/TRAINSBulkExport/TRAINSBulkExportQueryDefination.aspx?Page=TRAINSBulkExport. There are a lot of countries missing from the "including AVE" option, although they are included in the non "including AVE" option. I think that countries not included in "including AVE" have tariff rates only if they are in ad valorem format as reported by the country. That is, the World Bank hasn't converted these countries' tariffs from non-ad valorem to ad valorem.*
- *The tariff data is at the six-digit HS classification, as a result, two-digit and four-digit trade data is not included.*

Tariff Data Relative to Quantity Trade Gap

Combinations of year, product, and country pairs in the tariff data relative to combinations in the Comtrade data that aren't missing when subtracting reported import netweight from reported export netweight.

```
load(paste(DataPath,"Analysis Data/hs12_all_tariffs_qty.Rda", sep = "/"))
hs12_all_tariffs <- hs12_all_tariffs[, .(Year, ProductCode, Importer, Exporter)]

load(paste(DataPath,"Analysis Data/hs12_qty.Rda", sep = "/"))
hs12_qty <- hs12_qty[, .(Period, `Commodity Code`, Importer, Exporter)]

#For each year, how many product x o-d pairs in tariff data / trade product x o-d pairs?

product_year <- hs12_all_tariffs[, uniqueN(ProductCode), by=Year]
product_year <- rename(product_year, Products_tariffs = V1)

pair_year <- unique(setDT(hs12_all_tariffs), by = c("Importer", "Exporter", "Year"))
pair_year <- pair_year[, .N, by=Year]
pair_year <- rename(pair_year, Pairs_tariffs = N)

year_coverage <- merge(product_year, pair_year)

product_year_trade <- hs12_qty[, uniqueN(`Commodity Code`), by=Period]
product_year_trade <- rename(product_year_trade, Products_trade = V1)

pair_year_trade <- unique(setDT(hs12_qty), by = c("Importer", "Exporter", "Period"))
pair_year_trade <- pair_year_trade[, .N, by=Period]
pair_year_trade <- rename(pair_year_trade, Pairs_trade = N)

year_coverage_trade <- merge(product_year_trade, pair_year_trade)

year_coverage <- merge(year_coverage, year_coverage_trade, by.x = c("Year"), by.y = c("Period"), all = T)

year_coverage$Coverage <- (year_coverage$Products_tariffs*year_coverage$Pairs_tariffs)/
  (year_coverage$Products_trade*year_coverage$Pairs_trade)

year_coverage[is.na(year_coverage)] <- 0
```

```
pander(year_coverage)
```

Year	Products_tariffs	Pairs_tariffs	Products_trade	Pairs_trade	Coverage
2012	5196	4580	6420	7396	0.5012
2013	5197	6108	6423	9950	0.4967
2014	5194	6731	6420	11238	0.4846
2015	5191	7371	6417	12206	0.4885
2016	0	0	6413	6531	0

```
rm(pair_year, pair_year_trade, product_year, product_year_trade, year_coverage, year_coverage_trade)

#For each product, how many year x o-d pairs / all possible year x o-d pairs?

year_product <- hs12_all_tariffs[, uniqueN(`Year`), by=ProductCode]
year_product <- rename(year_product, Years_tariffs = V1)

pair_product <- unique(setDT(hs12_all_tariffs), by = c("Importer", "Exporter", "ProductCode"))
pair_product <- pair_product[, .N, by= .(ProductCode)]
pair_product <- rename(pair_product, Pairs_tariffs = N)

product_coverage <- merge(year_product, pair_product)

year_product_trade <- hs12_qty[, uniqueN(`Period`), by=`Commodity Code`]
year_product_trade <- rename(year_product_trade, Years_trade = V1)

pair_product_trade <- unique(setDT(hs12_qty), by = c("Importer", "Exporter", "Commodity Code"))
pair_product_trade <- pair_product_trade[, .N, by = .(`Commodity Code`)]
pair_product_trade <- rename(pair_product_trade, Pairs_trade = N)

product_coverage_trade <- merge(year_product_trade, pair_product_trade)

product_coverage <- merge(product_coverage, product_coverage_trade,
                          by.x = c("ProductCode"), by.y = c("Commodity Code"), all = T)

product_coverage$Coverage <- (product_coverage$Years_tariffs*product_coverage$Pairs_tariffs)/
  (product_coverage$Years_trade*product_coverage$Pairs_trade)

product_coverage[is.na(product_coverage)] <- 0

pander(product_coverage[order(Coverage)][1:10])
```

ProductCode	Years_tariffs	Pairs_tariffs	Years_trade	Pairs_trade	Coverage
0101	0	0	5	757	0
0102	0	0	5	705	0
0103	0	0	5	417	0
0104	0	0	5	367	0
0105	0	0	5	801	0
0106	0	0	5	2131	0
0201	0	0	5	1175	0
0202	0	0	5	1510	0
0203	0	0	5	1430	0
0204	0	0	5	925	0

```
pander(product_coverage[order(-Coverage)][1:10])
```

ProductCode	Years_tariffs	Pairs_tariffs	Years_trade	Pairs_trade	Coverage
020830	4	6	4	6	1
030195	3	3	3	3	1
292512	1	1	1	1	1
293341	4	4	4	4	1
811213	3	8	3	8	1
811252	1	1	1	1	1
890130	3	1	3	1	1
440341	4	7	4	8	0.875
293951	3	5	3	6	0.8333
030283	4	14	5	14	0.8

```
rm(pair_product, pair_product_trade, year_product,
   year_product_trade, product_coverage, product_coverage_trade)

#For each o-d pair, how many year x product / all possible year x product?

product_pair <- hs12_all_tariffs[, uniqueN(ProductCode), by = c("Importer", "Exporter")]
product_pair <- rename(product_pair, Products_tariffs = V1)

year_pair <- hs12_all_tariffs[, uniqueN(`Year`), by = c("Importer", "Exporter")]
year_pair <- rename(year_pair, Years_tariffs = V1)

pair_coverage <- merge(product_pair, year_pair, by = c("Importer", "Exporter"))

product_pair_trade <- hs12_qty[, uniqueN(`Commodity Code`), by = c("Importer", "Exporter")]
product_pair_trade <- rename(product_pair_trade, Products_trade = V1)

year_pair_trade <- hs12_qty[, uniqueN(`Period`), by = c("Importer", "Exporter")]
year_pair_trade <- rename(year_pair_trade, Years_trade = V1)

pair_coverage_trade <- merge(product_pair_trade, year_pair_trade)

pair_coverage <- merge(pair_coverage, pair_coverage_trade, all = T)

pair_coverage$Coverage <- (pair_coverage$Products_tariffs*pair_coverage$Years_tariffs)/
  (pair_coverage$Products_trade*pair_coverage$Years_trade)

pair_coverage[is.na(pair_coverage)] <- 0

pair_coverage$Exporter <- strtrim(pair_coverage$Exporter, 15)

pair_coverage[order(-Coverage)][1:10]
```

```
##        Importer        Exporter Products_tariffs Years_tariffs
##  1:  Bangladesh    Solomon Isds               1             1
##  2:      Canada           Niger               1             1
##  3:       Egypt Brunei Darussal               1             1
##  4:       Japan           Congo               1             1
##  5:     Belgium     Netherlands            4624             4
##  6: Netherlands         Belgium            4565             4
##  7:     Austria         Germany            4739             4
##  8: Netherlands         Germany            4711             4
```

```
##  9:     Germany     Netherlands             4672             4
## 10:      France     Netherlands             4558             4
##     Products_trade Years_trade  Coverage
##  1:              1           1 1.0000000
##  2:              1           1 1.0000000
##  3:              1           1 1.0000000
##  4:              1           1 1.0000000
##  5:           5733           4 0.8065585
##  6:           5678           4 0.8039803
##  7:           5911           4 0.8017256
##  8:           5892           4 0.7995587
##  9:           5847           4 0.7990422
## 10:           5729           4 0.7956013
```

```
rm(product_pair, product_pair_trade, year_pair, year_pair_trade, pair_coverage, pair_coverage_trade)
```

The next section looks at the number of product x year combinations for each importer in the tariff data relative to the trade data.

```
tariffs <- hs12_all_tariffs[, .N, by = "Importer"]
tariffs <- rename(tariffs, "Tariffs" = "N")

trade <- hs12_qty[, .N, by = "Importer"]
trade <- rename(trade, "Trade" = "N")

matches <- merge(tariffs, trade, by = c("Importer"), all = T)

matches[is.na(matches)] <- 0
matches$Share_covered <- matches$Tariffs / matches$Trade

pander(matches[order(-Share_covered)][1:10])
```

Importer	Tariffs	Trade	Share_covered
Austria	246703	345939	0.7131
Netherlands	317359	445774	0.7119
Finland	194294	274456	0.7079
Slovenia	173152	246961	0.7011
United Arab Emirates	251856	359334	0.7009
Ukraine	79503	115692	0.6872
Nepal	22203	32684	0.6793
Bangladesh	55277	81989	0.6742
Kuwait	99613	147789	0.674
Pakistan	84610	126147	0.6707

No country has tariffs for more than 72% of the trade data. This is mostly because there is no tariff data for year 2016, for product codes at the 2- or 4-digit level, and also for product code 999999, which is "Commodities not specified according to kind." The next section repeats the above but removing these conditions from the trade data.

```
trade <- hs12_qty[Period!=2016 & `Commodity Code`!="999999", ]
trade <- subset(trade, nchar(`Commodity Code`) > 4)

trade <- trade[, .N, by = "Importer"]
trade <- rename(trade, "Trade" = "N")

matches <- merge(tariffs, trade, by = c("Importer"), all = T)

matches[is.na(matches)] <- 0
```

```r
matches$Share_covered <- matches$Tariffs / matches$Trade

matches <- matches[order(-Share_covered)]
print(matches, nrow=144)
```

```
##                            Importer Tariffs  Trade Share_covered
##   1:                       Australia  229913 229913     1.0000000
##   2:                         Austria  246703 246703     1.0000000
##   3:                         Bahrain   71038  71038     1.0000000
##   4:                      Bangladesh   55277  55277     1.0000000
##   5:                         Belarus  113480 113480     1.0000000
##   6:                         Belgium  276139 276139     1.0000000
##   7:                         Bermuda   10818  10818     1.0000000
##   8:              Bosnia Herzegovina  114982 114982     1.0000000
##   9:                        Botswana   22097  22097     1.0000000
##  10:                        Bulgaria  179545 179545     1.0000000
##  11:                         Burundi    5431   5431     1.0000000
##  12:                          Canada  219292 219292     1.0000000
##  13:                         Croatia  173786 173786     1.0000000
##  14:                          Cyprus   97867  97867     1.0000000
##  15:                         Czechia  240268 240268     1.0000000
##  16:                         Denmark  223030 223030     1.0000000
##  17:                           Egypt   57755  57755     1.0000000
##  18:                         Estonia  151379 151379     1.0000000
##  19:                         Finland  194294 194294     1.0000000
##  20:                          France  374480 374480     1.0000000
##  21:                         Germany  447019 447019     1.0000000
##  22:                          Greece  183752 183752     1.0000000
##  23:                         Hungary  203784 203784     1.0000000
##  24:                         Iceland   85505  85505     1.0000000
##  25:                         Ireland  140087 140087     1.0000000
##  26:                          Israel  132338 132338     1.0000000
##  27:                           Italy  350310 350310     1.0000000
##  28:                      Kazakhstan  131090 131090     1.0000000
##  29:                          Kuwait   99613  99613     1.0000000
##  30:                      Kyrgyzstan   12421  12421     1.0000000
##  31:                          Latvia  162073 162073     1.0000000
##  32:                       Lithuania  182453 182453     1.0000000
##  33:                      Luxembourg  105907 105907     1.0000000
##  34:                           Malta   80185  80185     1.0000000
##  35:                       Mauritius   67823  67823     1.0000000
##  36:                      Montenegro   67041  67041     1.0000000
##  37:                         Namibia   39284  39284     1.0000000
##  38:                           Nepal   22203  22203     1.0000000
##  39:                     Netherlands  317359 317359     1.0000000
##  40:                     New Zealand  156775 156775     1.0000000
##  41:                          Norway  215622 215622     1.0000000
##  42:                        Pakistan   84610  84610     1.0000000
##  43:                Papua New Guinea    2873   2873     1.0000000
##  44:                          Poland  266646 266646     1.0000000
##  45:                        Portugal  176458 176458     1.0000000
##  46:                           Qatar   95461  95461     1.0000000
##  47:                  Rep. of Korea  218811 218811     1.0000000
##  48:                 Rep. of Moldova   21229  21229     1.0000000
##  49:                         Romania  234522 234522     1.0000000
##  50:              Russian Federation  264620 264620     1.0000000
##  51:                          Rwanda   18329  18329     1.0000000
```

```
##  52:                           Samoa    9558   9558     1.0000000
##  53:                       Singapore  180877 180877     1.0000000
##  54:                        Slovakia  170487 170487     1.0000000
##  55:                        Slovenia  173152 173152     1.0000000
##  56:                    Solomon Isds    2894   2894     1.0000000
##  57:                    South Africa  198568 198568     1.0000000
##  58:                           Spain  297416 297416     1.0000000
##  59:              State of Palestine    2849   2849     1.0000000
##  60:                          Sweden  235450 235450     1.0000000
##  61:                     Switzerland  284411 284411     1.0000000
##  62:              TFYR of Macedonia   97530  97530     1.0000000
##  63:                             USA  342470 342470     1.0000000
##  64:                          Uganda   44493  44493     1.0000000
##  65:                         Ukraine   79503  79503     1.0000000
##  66:           United Arab Emirates  251856 251856     1.0000000
##  67:                 United Kingdom  340120 340120     1.0000000
##  68:        United Rep. of Tanzania   57906  57906     1.0000000
##  69:                        Thailand  152554 192110     0.7940971
##  70:                      Azerbaijan   50772  68017     0.7464604
##  71:                           Japan  152625 205613     0.7422926
##  72:                    Saudi Arabia   83805 114419     0.7324395
##  73:                         Ecuador   50968  77075     0.6612780
##  74:                           Palau    1704   3051     0.5585054
##  75:                           China  144416 265445     0.5440524
##  76:                         Georgia   51440  99751     0.5156841
##  77:                          Jordan   37359  76453     0.4886532
##  78:                            Fiji   13880  32866     0.4223209
##  79:                           Aruba    2469   6344     0.3891866
##  80:                         Lebanon   28463  80414     0.3539558
##  81:                       Sri Lanka   22654  65914     0.3436903
##  82:                         Armenia   12612  37966     0.3321920
##  83:              Brunei Darussalam   10323  31078     0.3321642
##  84:                           India   52357 160734     0.3257369
##  85:                        Malaysia   44443 136937     0.3245507
##  86:                        Zimbabwe   10103  31692     0.3187871
##  87:                            Oman   23902  88607     0.2697530
##  88:                          Turkey   56533 216000     0.2617269
##  89:                         Albania       0  49577     0.0000000
##  90:                         Algeria       0  95530     0.0000000
##  91:                         Andorra       0  19721     0.0000000
##  92:                          Angola       0  56849     0.0000000
##  93:                       Argentina       0 110693     0.0000000
##  94:                         Bahamas       0  12543     0.0000000
##  95:                          Belize       0   9829     0.0000000
##  96:                           Benin       0   5034     0.0000000
##  97:                          Bhutan       0    669     0.0000000
##  98: Bolivia (Plurinational State of)       0  72978     0.0000000
##  99:                          Brazil       0 186853     0.0000000
## 100:                    Burkina Faso       0   6066     0.0000000
## 101:                      Cabo Verde       0  17422     0.0000000
## 102:                        Cambodia       0  33355     0.0000000
## 103:                        Cameroon       0  42171     0.0000000
## 104:                           Chile       0 156209     0.0000000
## 105:           China, Hong Kong SAR       0 149684     0.0000000
## 106:               China, Macao SAR       0  18280     0.0000000
## 107:                        Colombia       0 136861     0.0000000
## 108:                           Congo       0  11660     0.0000000
```

```
## 109:                      Costa Rica       0  97207     0.0000000
## 110:                   CÃ´te d'Ivoire       0  17299     0.0000000
## 111:                   Dominican Rep.      0  68547     0.0000000
## 112:                      El Salvador      0  71591     0.0000000
## 113:                         Ethiopia      0  28198     0.0000000
## 114:                 French Polynesia      0  36320     0.0000000
## 115:                            Ghana      0  22621     0.0000000
## 116:                        Greenland      0  14374     0.0000000
## 117:                        Guatemala      0  93016     0.0000000
## 118:                           Guinea      0  18727     0.0000000
## 119:                         Honduras      0  30669     0.0000000
## 120:                        Indonesia      0 160566     0.0000000
## 121:                          Jamaica      0  17939     0.0000000
## 122:                       Madagascar      0  40881     0.0000000
## 123:                           Malawi      0  14260     0.0000000
## 124:                         Maldives      0  36606     0.0000000
## 125:                           Mexico      0 185323     0.0000000
## 126:                         Mongolia      0  33543     0.0000000
## 127:                          Morocco      0  32626     0.0000000
## 128:                    New Caledonia      0  21777     0.0000000
## 129:                        Nicaragua      0  53826     0.0000000
## 130:                            Niger      0   4239     0.0000000
## 131:                  Other Asia, nes      0 134608     0.0000000
## 132:                           Panama      0  83796     0.0000000
## 133:                         Paraguay      0  64581     0.0000000
## 134:                             Peru      0 126194     0.0000000
## 135:                          Senegal      0  13937     0.0000000
## 136:                           Serbia      0 155138     0.0000000
## 137:                            Sudan      0   6378     0.0000000
## 138:                             Togo      0   4889     0.0000000
## 139:                            Tonga      0   2584     0.0000000
## 140:                          Tunisia      0  97583     0.0000000
## 141:                          Uruguay      0  90376     0.0000000
## 142:                         Viet Nam      0  99831     0.0000000
## 143:                            Yemen      0  23815     0.0000000
## 144:                           Zambia      0  48232     0.0000000
##                              Importer Tariffs  Trade Share_covered
```

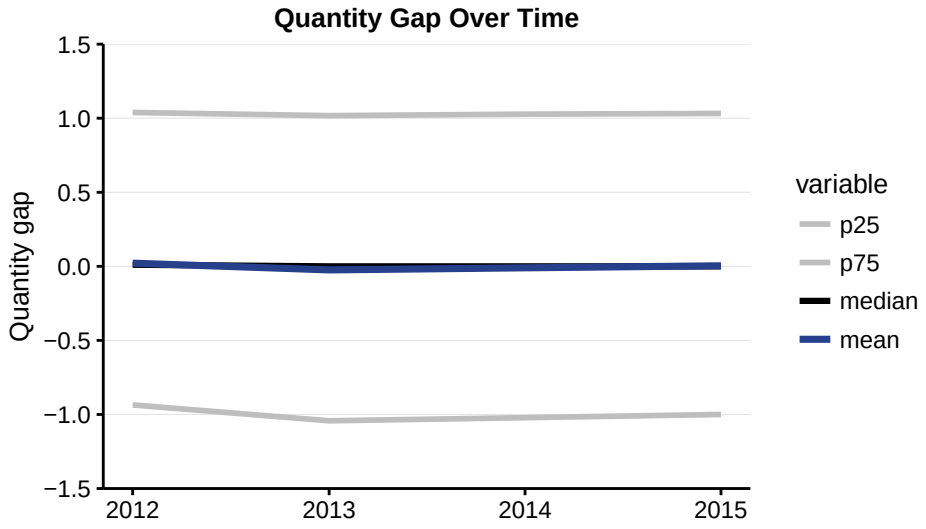
```
rm(hs12_all_tariffs, hs12_qty, matches, tariffs, trade)
```

Quantity Trade Gap, Limited to Data with Corresponding Tariff Rates

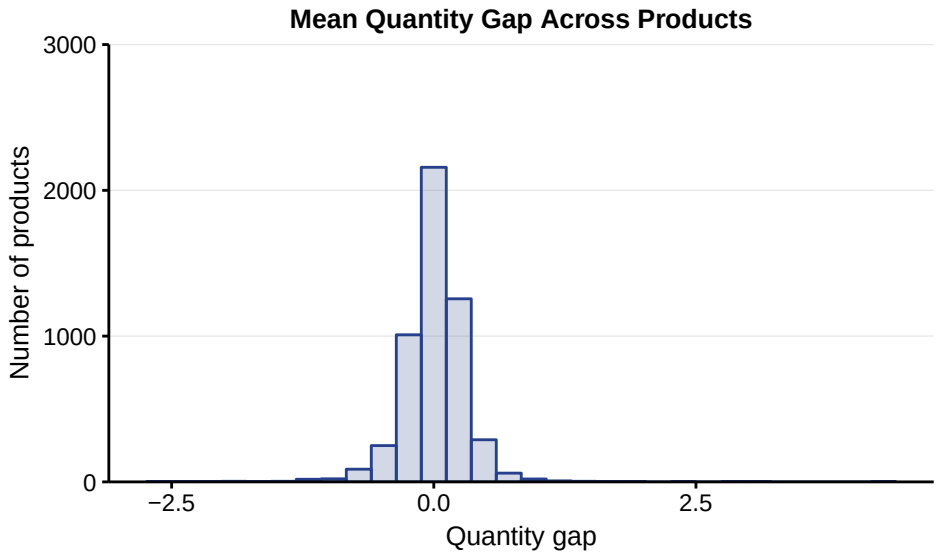
The following figures are the same as the previous summary files on the quantity evasion gap (log exports - log imports), but limited to data with tariff rates.

```
#How has the trade gap changed over time?

load(paste(DataPath,"Analysis Data/hs12_all_tariffs_qty.Rda", sep = "/"))

hs12_all_tariffs <- hs12_all_tariffs[, .(Year, ProductCode, Importer, Exporter, Qty_log_gap)]

hs12_all_tariffs$Year <- as.Date(hs12_all_tariffs$Year, "%Y")
hs12_all_tariffs$Year <- floor_date(hs12_all_tariffs$Year,"year")

Years <- hs12_all_tariffs[, .(mean = as.double(mean(Qty_log_gap)),
                              median = as.double(median(Qty_log_gap)),
```

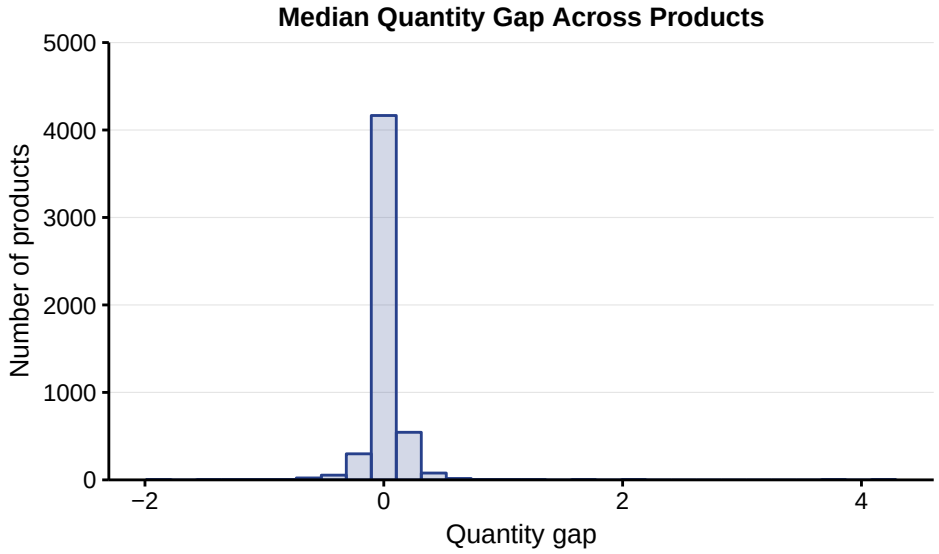
```
                              p25 = as.double(quantile(Qty_log_gap,.25)),
                              p75 = as.double(quantile(Qty_log_gap,.75))
),
by=Year]

Years <- melt(Years, id = 'Year')
Years$variable <- factor(Years$variable, levels = c("p25","p75","median","mean"))

ggplot(data=Years ) +
  geom_line(data=Years, aes(x = Year, y = value, colour = variable, size=variable)) +
  scale_colour_manual(values=c("grey","grey","black","royalblue4")) +
  background_grid(major = 'y', minor = "none") +
  scale_size_manual(values = c(1,1,1.1,1.25)) +
  scale_y_continuous(expand = c(0, 0), limits = c(-1.5,1.5), minor_breaks = NULL) +
  xlab(label = "") +
  ylab(label = "Quantity gap") +
  labs(title="Quantity Gap Over Time")
```



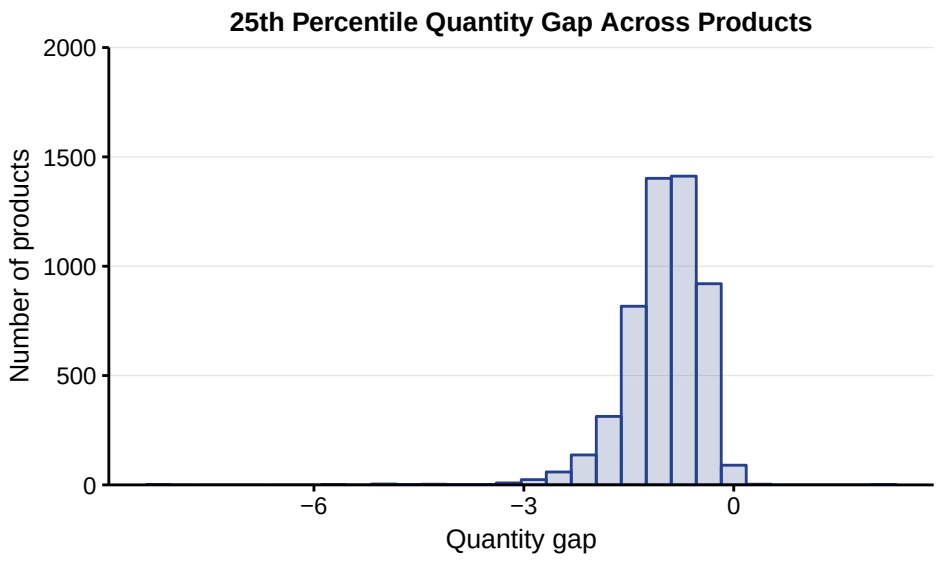

```
#Across products?
products <- hs12_all_tariffs[, .(mean = as.double(mean(Qty_log_gap)),
                                 median = as.double(median(Qty_log_gap)),
                                 p25 = as.double(quantile(Qty_log_gap,.25)),
                                 p75 = as.double(quantile(Qty_log_gap,.75))
),
by= ProductCode]

ggplot(data=products, aes(mean)) +
  geom_histogram(col="royalblue4",
                 fill="royalblue4",
                 alpha=.2) +
  background_grid(major = 'y', minor = "none") +
  scale_y_continuous(expand = c(0, 0), limits = c(0, 3000))  +
  labs(title="Mean Quantity Gap Across Products") +
  labs(x="Quantity gap", y="Number of products")
```



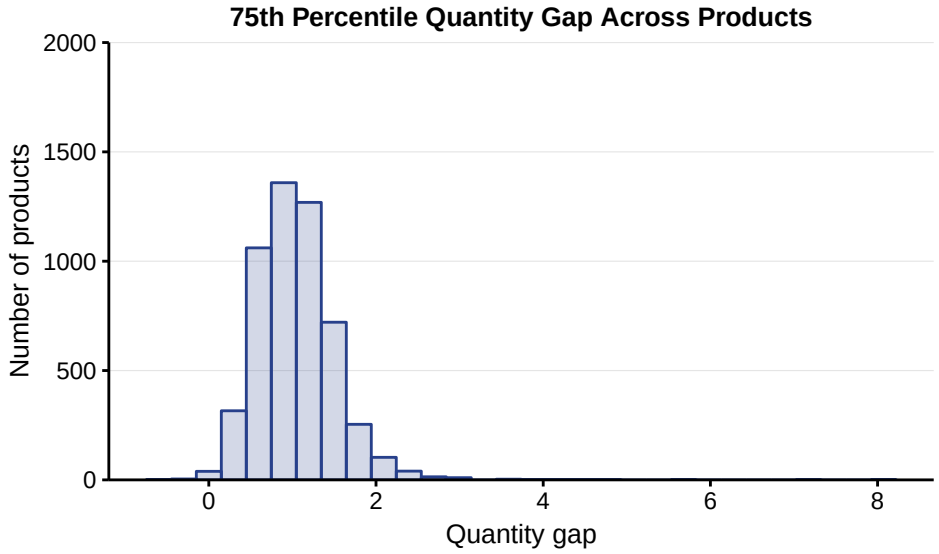

```
ggplot(data=products, aes(median)) +
  geom_histogram(col="royalblue4",
                 fill="royalblue4",
                 alpha=.2) +
  background_grid(major = 'y', minor = "none") +
  scale_y_continuous(expand = c(0, 0), limits = c(0, 5000)) +
  labs(title="Median Quantity Gap Across Products") +
  labs(x="Quantity gap", y="Number of products")
```




```
ggplot(data=products, aes(p25)) +
  geom_histogram(col="royalblue4",
                 fill="royalblue4",
                 alpha=.2) +
  background_grid(major = 'y', minor = "none") +
  scale_y_continuous(expand = c(0, 0), limits = c(0,2000),  minor_breaks = NULL) +
  labs(title="25th Percentile Quantity Gap Across Products") +
  labs(x="Quantity gap", y="Number of products")
```

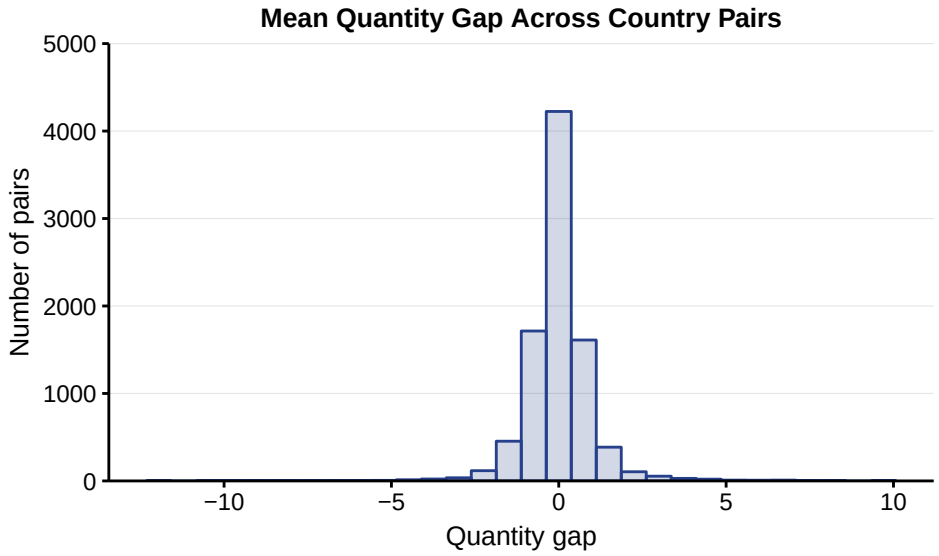



```
ggplot(data=products, aes(p75)) +
  geom_histogram(col="royalblue4",
                 fill="royalblue4",
                 alpha=.2) +
  background_grid(major = 'y', minor = "none") +
  scale_y_continuous(expand = c(0, 0),  limits = c(0, 2000), minor_breaks = NULL) +
  labs(title="75th Percentile Quantity Gap Across Products") +
  labs(x="Quantity gap", y="Number of products")
```

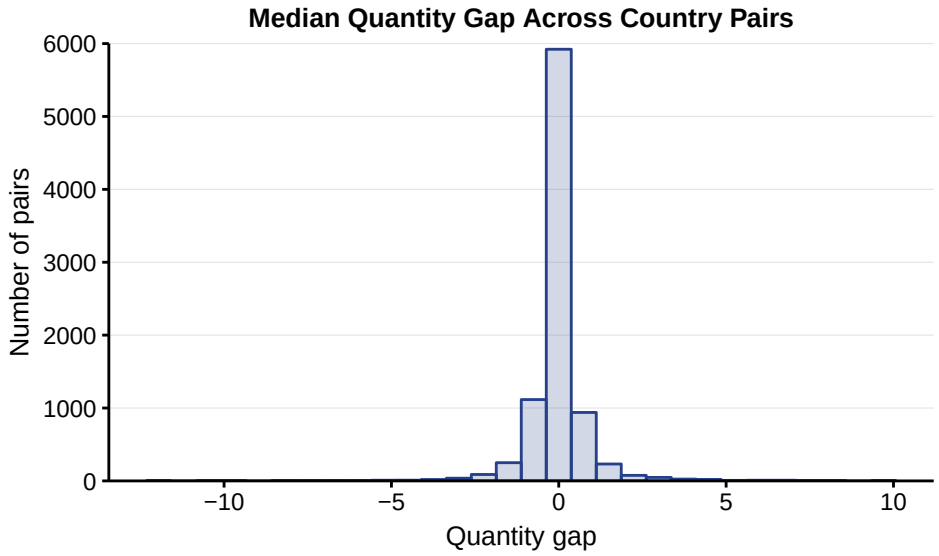



```
#Across countries?
countries <- hs12_all_tariffs[, .(mean = as.double(mean(Qty_log_gap)),
                                  median = as.double(median(Qty_log_gap)),
                                  p25 = as.double(quantile(Qty_log_gap,.25)),
                                  p75 = as.double(quantile(Qty_log_gap,.75))
),
by= c("Importer", "Exporter")]

ggplot(data=countries, aes(mean)) +
  geom_histogram(col="royalblue4",
                 fill="royalblue4",
```

```
                alpha=.2) +
  background_grid(major = 'y', minor = "none") +
  scale_y_continuous(expand = c(0, 0), limits = c(0, 5000),  minor_breaks = NULL) +
  labs(title="Mean Quantity Gap Across Country Pairs") +
  labs(x="Quantity gap", y="Number of pairs")
```

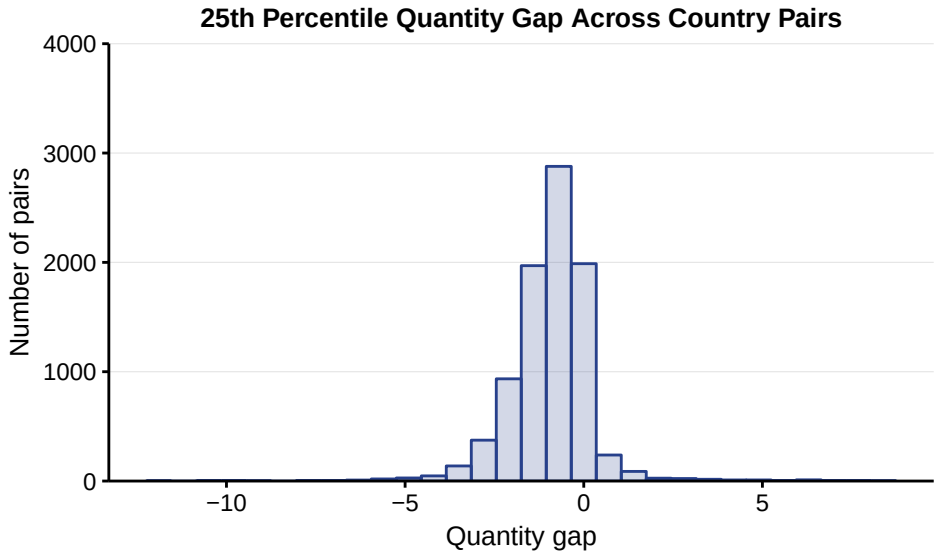



```
ggplot(data=countries, aes(median)) +
  geom_histogram(col="royalblue4",
                 fill="royalblue4",
                 alpha=.2) +
  background_grid(major = 'y', minor = "none") +
  scale_y_continuous(expand = c(0, 0), limits = c(0, 6000), minor_breaks = NULL) +
  labs(title="Median Quantity Gap Across Country Pairs") +
  labs(x="Quantity gap", y="Number of pairs")
```

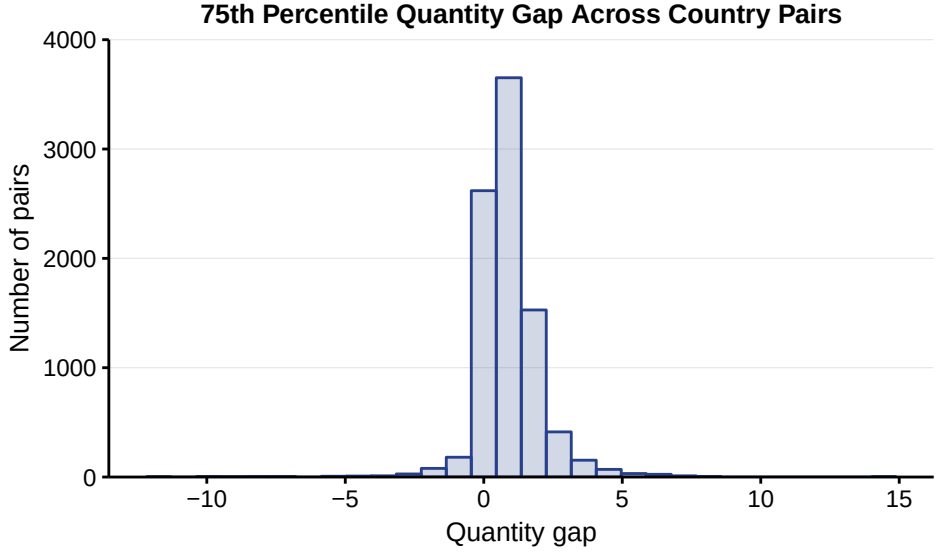



```
ggplot(data=countries, aes(p25)) +
  geom_histogram(col="royalblue4",
                 fill="royalblue4",
                 alpha=.2) +
```

```
  background_grid(major = 'y', minor = "none") +
  scale_y_continuous(expand = c(0, 0), limits = c(0, 4000),  minor_breaks = NULL) +
  labs(title="25th Percentile Quantity Gap Across Country Pairs") +
  labs(x="Quantity gap", y="Number of pairs")
```

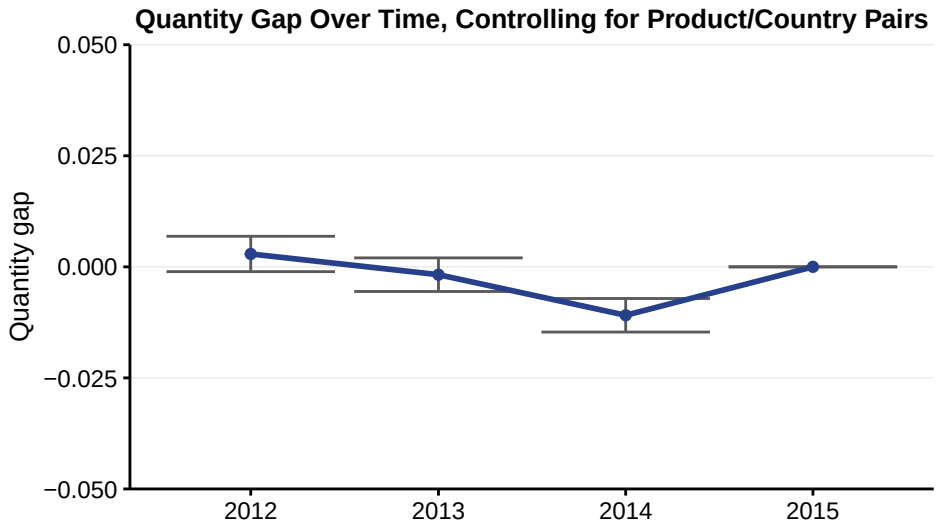



```
ggplot(data=countries, aes(p75)) +
  geom_histogram(col="royalblue4",
                 fill="royalblue4",
                 alpha=.2) +
  background_grid(major = 'y', minor = "none") +
  scale_y_continuous(expand = c(0, 0), limits = c(0, 4000),  minor_breaks = NULL) +
  labs(title="75th Percentile Quantity Gap Across Country Pairs") +
  labs(x="Quantity gap", y="Number of pairs")
```

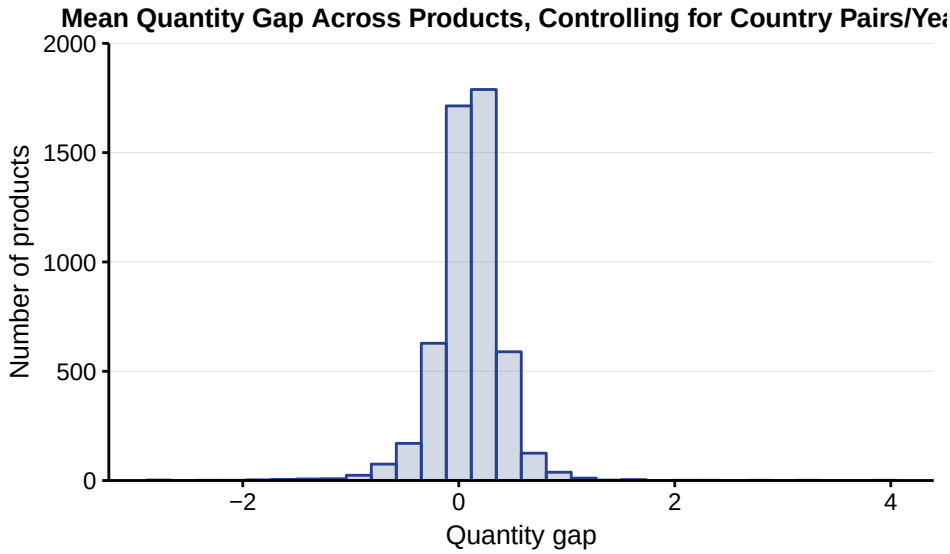



```
rm(Years, products, countries, hs12_all_tariffs)

#Regress trade gap on dummies and plot coefficients

load(paste(DataPath,"Analysis Data/hs12_all_tariffs_qty.Rda", sep = "/"))
```

```r
hs12_all_tariffs <- hs12_all_tariffs[, .(Year, ProductCode, `Reporter_ISO_N`, `Partner Code`, Qty_log_gap)]

hs12_all_tariffs$Year <- as.Date(hs12_all_tariffs$Year, "%Y")
hs12_all_tariffs$Year <- floor_date(hs12_all_tariffs$Year,"year")

hs12_all_tariffs$Year.f <- factor(hs12_all_tariffs$Year)
hs12_all_tariffs$Products.f <- factor(hs12_all_tariffs$ProductCode)

hs12_all_tariffs$Importer.f <- factor(hs12_all_tariffs$`Reporter_ISO_N`)
hs12_all_tariffs$Exporter.f <- factor(hs12_all_tariffs$`Partner Code`)
hs12_all_tariffs$Pairs.f <- with(hs12_all_tariffs, interaction(Importer.f, Exporter.f))

hs12_all_tariffs <- hs12_all_tariffs[, .(Year, Qty_log_gap, Year.f, Products.f, Pairs.f)]

reg <- felm(Qty_log_gap ~ 1 | Year.f + Products.f + Pairs.f,
            data = hs12_all_tariffs,
            exactDOF = FALSE,
            keepX = FALSE,
            keepCX = FALSE)

fes <- getfe(reg,
             se=TRUE,
             bN = 50
)

Yearfes <- subset(fes,fe == "Year.f")

Yearfes$ci_ub <- Yearfes$effect + (1.96 * Yearfes$se)
Yearfes$ci_lb <- Yearfes$effect - (1.96 * Yearfes$se)
Yearfes <- merge(Yearfes,unique(hs12_all_tariffs[,list(Year,Year.f)]),by.x = "idx",by.y="Year.f")
Yearfes <- rename(Yearfes, Year = Year)

ggplot(data = Yearfes, aes(Year,effect)) +
  geom_errorbar(aes(ymin = ci_lb, ymax = ci_ub), color = "grey35") +
  geom_line(color = "royalblue4", size = 1) +
  geom_point(color = "royalblue4") +
  background_grid(major = 'y', minor = "none") +
  scale_y_continuous(expand = c(0, 0), limits = c(-.05,.05), minor_breaks = NULL) +
  xlab(label = "") +
  ylab(label = "Quantity gap") +
  labs(title = "Quantity Gap Over Time, Controlling for Product/Country Pairs")
```

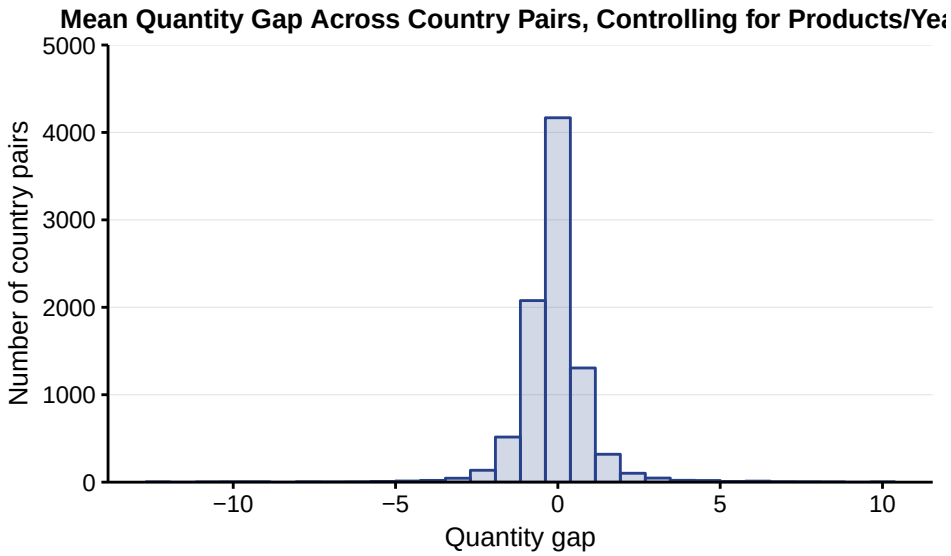



```
productfes <- subset(fes,fe == "Products.f")
productfes <- productfes[,c("effect","idx")]

ggplot(data=productfes, aes(effect)) +
  geom_histogram(col="royalblue4",
                 fill="royalblue4",
                 alpha=.2) +
  background_grid(major = 'y', minor = "none") +
  scale_y_continuous(expand = c(0, 0), limits = c(0,2000), minor_breaks = NULL) +
  labs(title="Mean Quantity Gap Across Products, Controlling for Country Pairs/Years") +
  labs(x="Quantity gap", y="Number of products")
```




```
pairfes <- subset(fes,fe == "Pairs.f")
pairfes <- pairfes[,c("effect","idx")]

ggplot(data=pairfes, aes(effect)) +
  geom_histogram(col="royalblue4",
                 fill="royalblue4",
                 alpha=.2) +
  background_grid(major = 'y', minor = "none") +
```

```
  scale_y_continuous(expand = c(0, 0), limits = c(0, 5000), minor_breaks = NULL) +
  labs(title="Mean Quantity Gap Across Country Pairs, Controlling for Products/Years") +
  labs(x="Quantity gap", y="Number of country pairs")
```



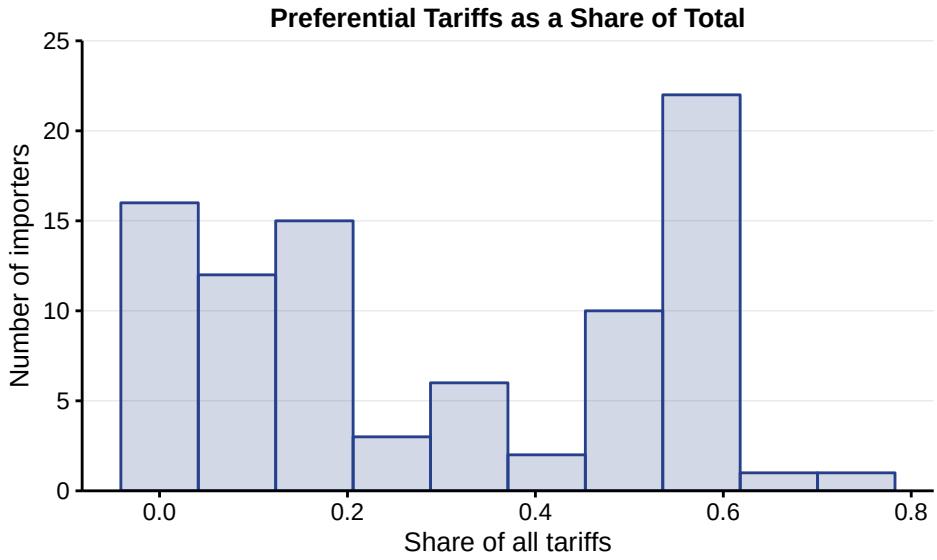

```
rm(fes, hs12_all_tariffs, pairfes, Yearfes, productfes, reg)
```

Tariff Data Summary

The following figure presents preferential tariffs as a share of each country's total tariffs, for data that matches with the quantity trade data.

```
load(paste(DataPath,"Analysis Data/hs12_all_tariffs_qty.Rda", sep = "/"))

hs12_all_tariffs <- hs12_all_tariffs[, .(Year, ProductCode, Importer, Exporter, pref)]

pref <- hs12_all_tariffs[pref==1, .N, by = c("Importer")]
pref <- rename(pref, "Pref" = "N")

mfn <- hs12_all_tariffs[is.na(pref), .N, by = c("Importer")]
mfn <- rename(mfn, "MFN" = "N")

tariffs <- merge(pref, mfn, by = c("Importer"), all = T)

tariffs[is.na(tariffs)] <- 0

tariffs$All <- tariffs$Pref + tariffs$MFN
tariffs$Share_pref <- tariffs$Pref / tariffs$All

ggplot(tariffs, aes(Share_pref)) +
  geom_histogram(col="royalblue4",
                 fill="royalblue4",
                 bins = 10,
                 alpha=.2) +
  background_grid(major = 'y', minor = "none") +
  scale_y_continuous(expand = c(0, 0), limits = c(0, 25), minor_breaks = NULL) +
  labs(title="Preferential Tariffs as a Share of Total") +
```

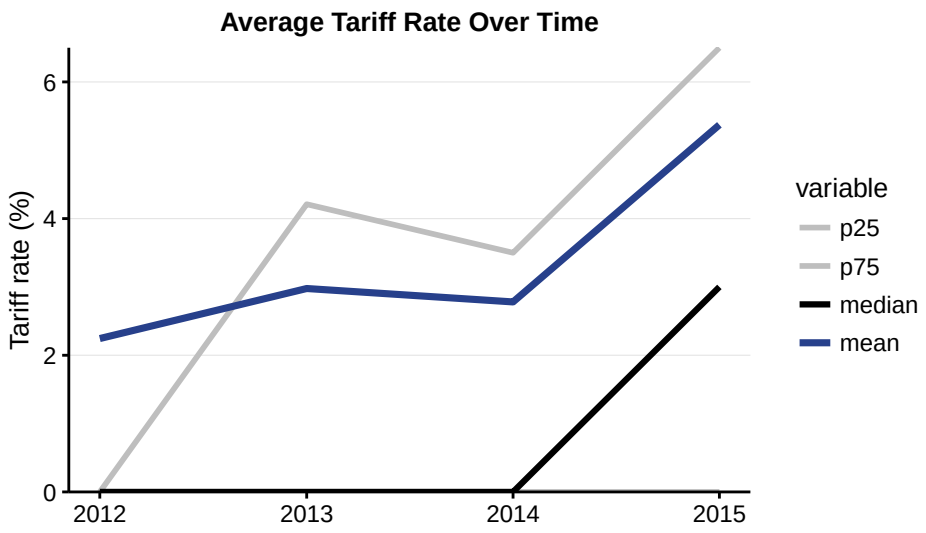
```
  labs(x="Share of all tariffs", y="Number of importers")
```



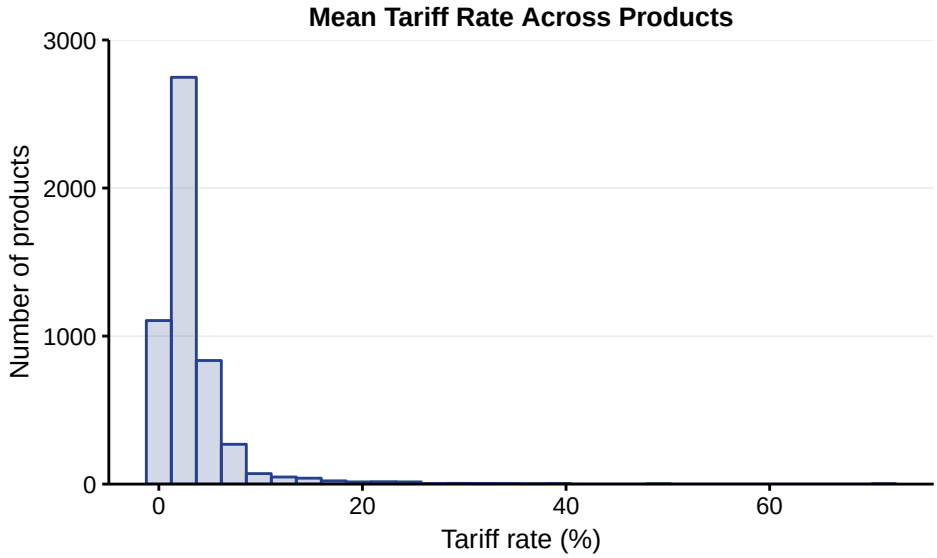

```
rm(hs12_all_tariffs, mfn, pref, tariffs)
```

The next section summarizes the "Simple Average" tariff rate as reported by WITS, for tariffs that match with the quantity trade data. The Simple Average is the average ad valorem tariff rate within each six-digit HS code. Each tariff is the most-favored nation rate unless there is a corresponding preferential tariff rate. There were some instances of multiple preferential tariff rates, in which case I took the lowest value if the average was different.

```
load(paste(DataPath,"Analysis Data/hs12_all_tariffs_qty.Rda", sep = "/"))

hs12_all_tariffs <- hs12_all_tariffs[, .(Year, ProductCode, Importer, Exporter, SimpleAverage)]
hs12_all_tariffs <- hs12_all_tariffs[!is.na(SimpleAverage)]

hs12_all_tariffs$Year <- as.Date(hs12_all_tariffs$Year, "%Y")
hs12_all_tariffs$Year <- floor_date(hs12_all_tariffs$Year,"year")

Years <- hs12_all_tariffs[, .(mean = as.double(mean(SimpleAverage)),
                              median = as.double(median(SimpleAverage)),
                              p25 = as.double(quantile(SimpleAverage,.25)),
                              p75 = as.double(quantile(SimpleAverage,.75))
), by=Year]

Years <- melt(Years, id = 'Year')
Years$variable <- factor(Years$variable, levels = c("p25","p75","median","mean"))

ggplot(data=Years ) +
  geom_line(data=Years, aes(x = Year, y = value, colour = variable, size=variable)) +
  scale_colour_manual(values=c("grey","grey","black","royalblue4")) +
  background_grid(major = 'y', minor = "none") +
  scale_size_manual(values = c(1,1,1.1,1.25)) +
  scale_y_continuous(expand = c(0, 0), minor_breaks = NULL) +
  xlab(label = "") +
  ylab(label = "Tariff rate (%)") +
  labs(title="Average Tariff Rate Over Time")
```

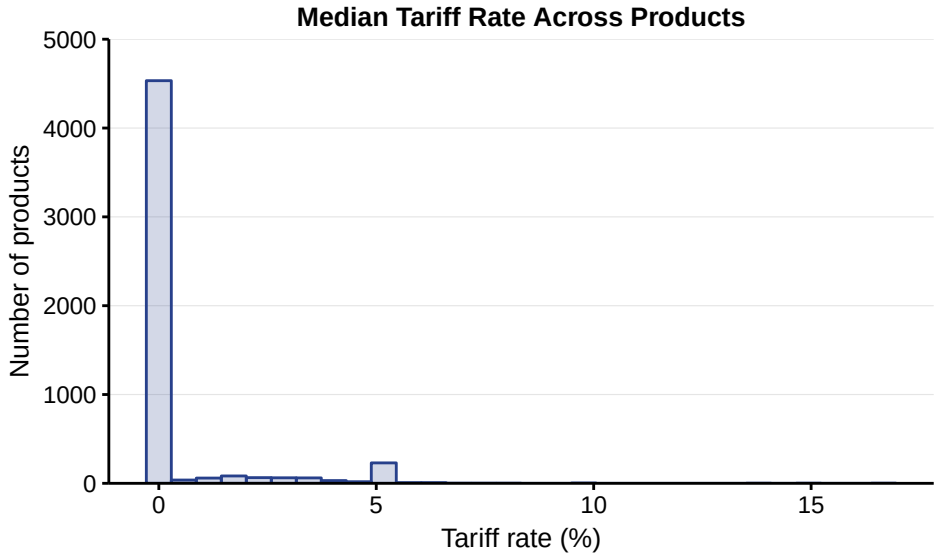



```
#Across products?
products <- hs12_all_tariffs[, .(mean = as.double(mean(SimpleAverage)),
                                 median = as.double(median(SimpleAverage)),
                                 p25 = as.double(quantile(SimpleAverage,.25)),
                                 p75 = as.double(quantile(SimpleAverage,.75))
),
by= ProductCode]

ggplot(data=products, aes(mean)) +
  geom_histogram(col="royalblue4",
                 fill="royalblue4",
                 alpha=.2) +
  background_grid(major = 'y', minor = "none") +
  scale_y_continuous(expand = c(0, 0), limits = c(0,3000))  +
  labs(title="Mean Tariff Rate Across Products") +
  labs(x="Tariff rate (%)", y="Number of products")
```

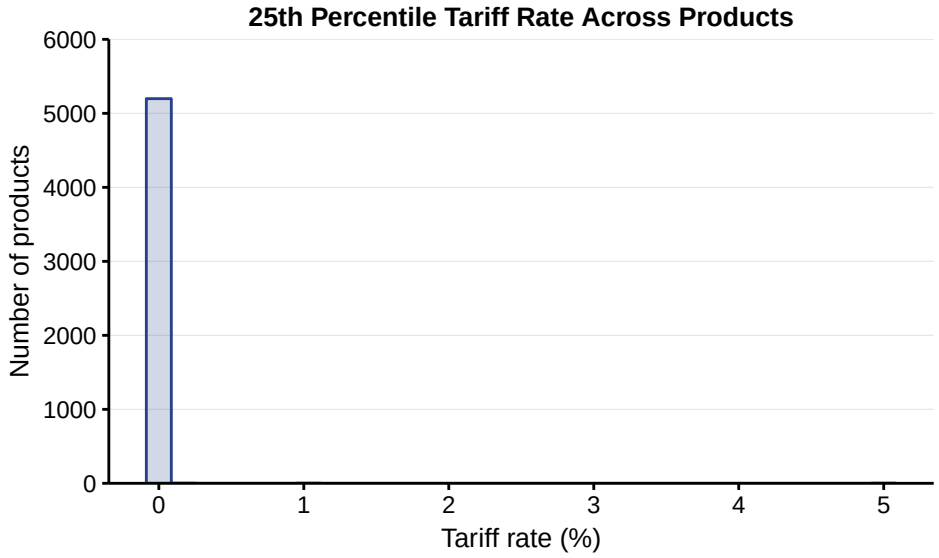



```
ggplot(data=products, aes(median)) +
  geom_histogram(col="royalblue4",
                 fill="royalblue4",
```

```
                alpha=.2) +
  background_grid(major = 'y', minor = "none") +
  scale_y_continuous(expand = c(0, 0), limits = c(0, 5000)) +
  labs(title="Median Tariff Rate Across Products") +
  labs(x="Tariff rate (%)", y="Number of products")
```

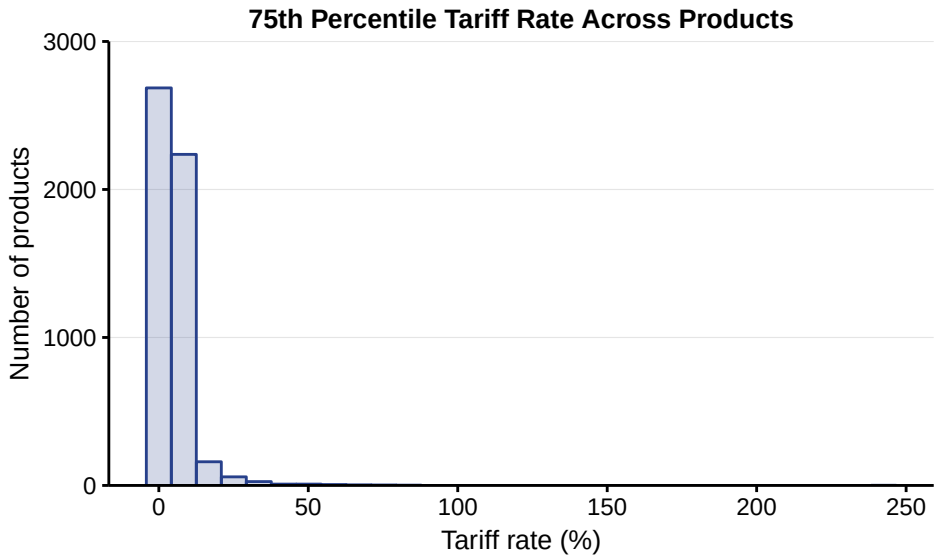



```
ggplot(data=products, aes(p25)) +
  geom_histogram(col="royalblue4",
                 fill="royalblue4",
                 alpha=.2) +
  background_grid(major = 'y', minor = "none") +
  scale_y_continuous(expand = c(0, 0), limits = c(0, 6000), minor_breaks = NULL) +
  labs(title="25th Percentile Tariff Rate Across Products") +
  labs(x="Tariff rate (%)", y="Number of products")
```

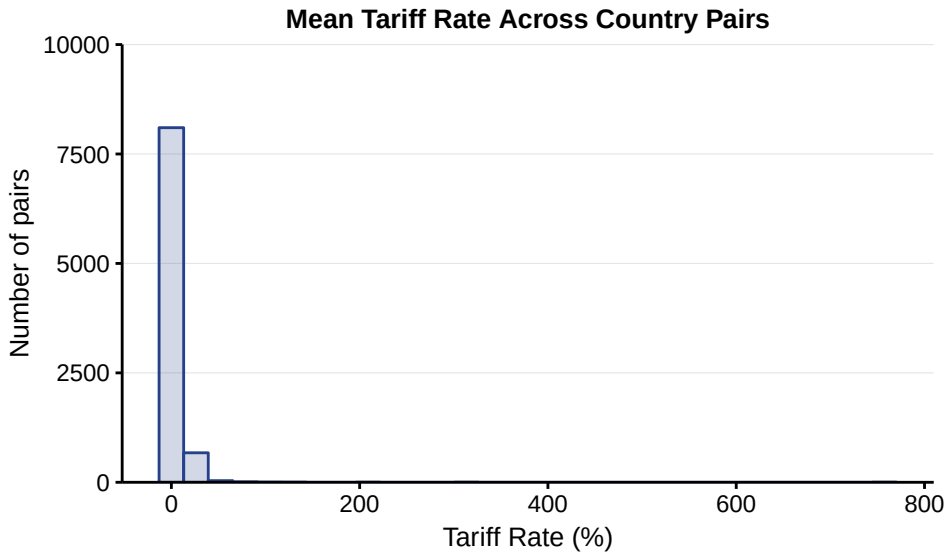



```
ggplot(data=products, aes(p75)) +
  geom_histogram(col="royalblue4",
                 fill="royalblue4",
                 alpha=.2) +
```

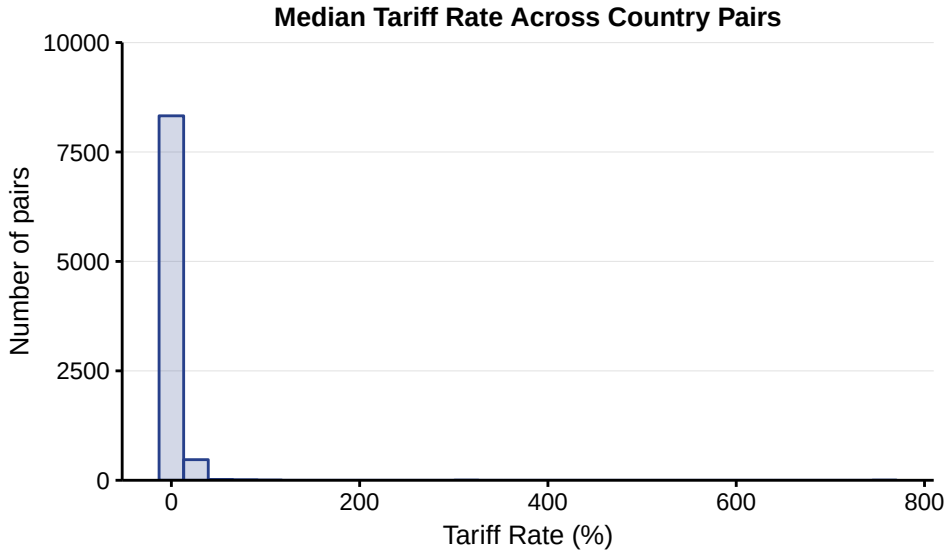
```
  background_grid(major = 'y', minor = "none") +
  scale_y_continuous(expand = c(0, 0), limits = c(0, 3000),  minor_breaks = NULL) +
  labs(title="75th Percentile Tariff Rate Across Products") +
  labs(x="Tariff rate (%)", y="Number of products")
```



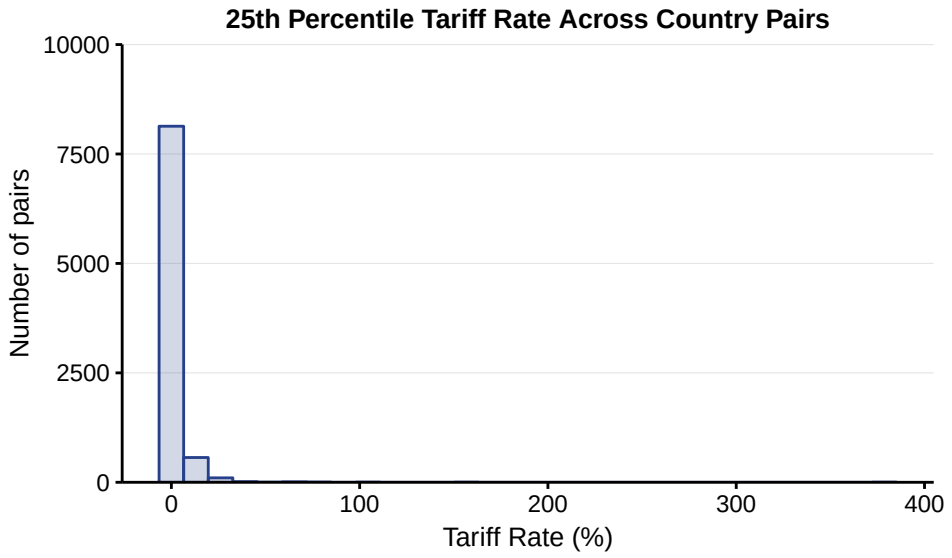

```
#Across countries?
countries <- hs12_all_tariffs[, .(mean = as.double(mean(SimpleAverage)),
                                  median = as.double(median(SimpleAverage)),
                                  p25 = as.double(quantile(SimpleAverage,.25)),
                                  p75 = as.double(quantile(SimpleAverage,.75))
),
by= c("Importer", "Exporter")]

ggplot(data=countries, aes(mean)) +
  geom_histogram(col="royalblue4",
                 fill="royalblue4",
                 alpha=.2) +
  background_grid(major = 'y', minor = "none") +
  scale_y_continuous(expand = c(0, 0), limits = c(0, 10000), minor_breaks = NULL) +
  labs(title="Mean Tariff Rate Across Country Pairs") +
  labs(x="Tariff Rate (%)", y="Number of pairs")
```



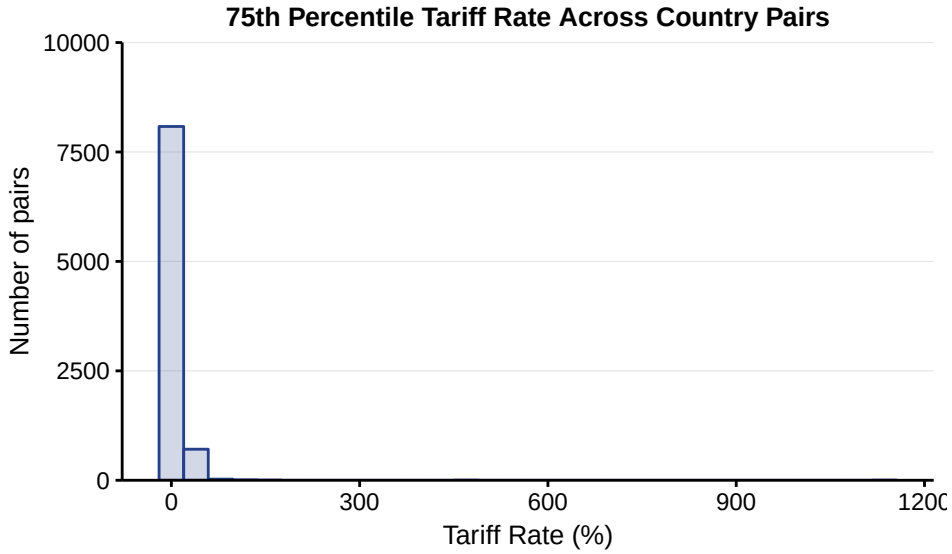

```
ggplot(data=countries, aes(median)) +
  geom_histogram(col="royalblue4",
                 fill="royalblue4",
                 alpha=.2) +
  background_grid(major = 'y', minor = "none") +
  scale_y_continuous(expand = c(0, 0), limits = c(0, 10000), minor_breaks = NULL) +
  labs(title="Median Tariff Rate Across Country Pairs") +
  labs(x="Tariff Rate (%)", y="Number of pairs")
```




```
ggplot(data=countries, aes(p25)) +
  geom_histogram(col="royalblue4",
                 fill="royalblue4",
                 alpha=.2) +
  background_grid(major = 'y', minor = "none") +
  scale_y_continuous(expand = c(0, 0), limits = c(0, 10000), minor_breaks = NULL) +
  labs(title="25th Percentile Tariff Rate Across Country Pairs") +
  labs(x="Tariff Rate (%)", y="Number of pairs")
```

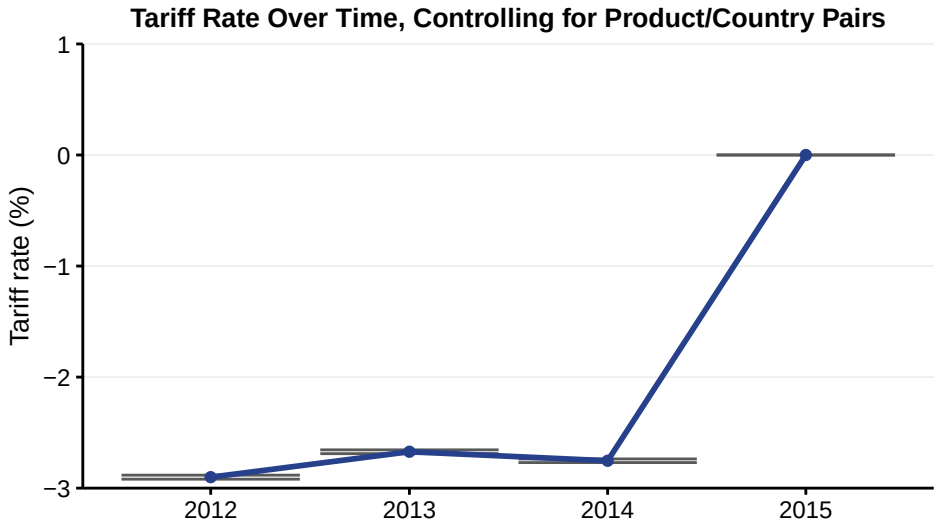



```
ggplot(data=countries, aes(p75)) +
  geom_histogram(col="royalblue4",
                 fill="royalblue4",
                 alpha=.2) +
  background_grid(major = 'y', minor = "none") +
  scale_y_continuous(expand = c(0, 0), limits = c(0, 10000), minor_breaks = NULL) +
  labs(title="75th Percentile Tariff Rate Across Country Pairs") +
  labs(x="Tariff Rate (%)", y="Number of pairs")
```

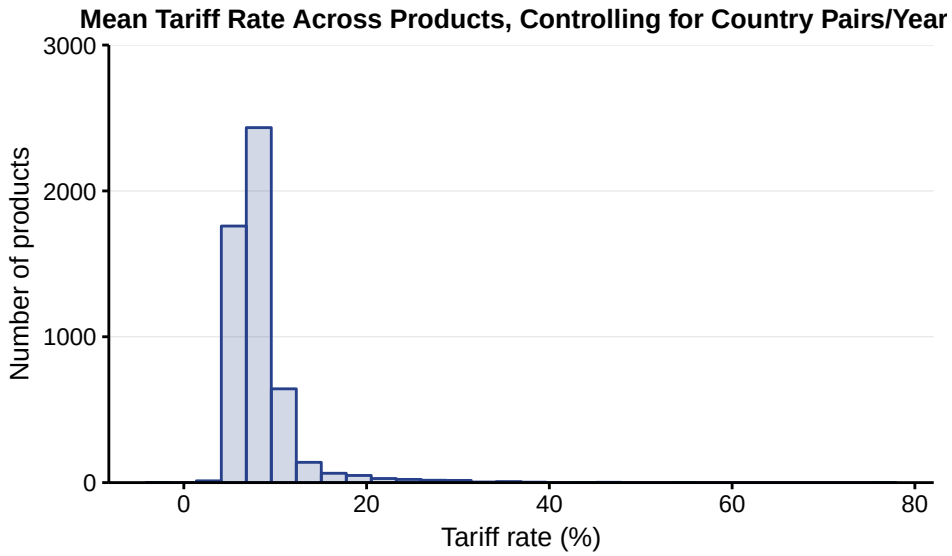



```
rm(Years, products, countries, hs12_all_tariffs)

#Regress trade gap on dummies and plot coefficients

load(paste(DataPath,"Analysis Data/hs12_all_tariffs_qty.Rda", sep = "/"))
hs12_all_tariffs <- hs12_all_tariffs[!is.na(SimpleAverage)]

hs12_all_tariffs <- hs12_all_tariffs[,
                    .(Year, ProductCode, `Reporter_ISO_N`, `Partner Code`, SimpleAverage)]

hs12_all_tariffs$Year <- as.Date(hs12_all_tariffs$Year, "%Y")
```

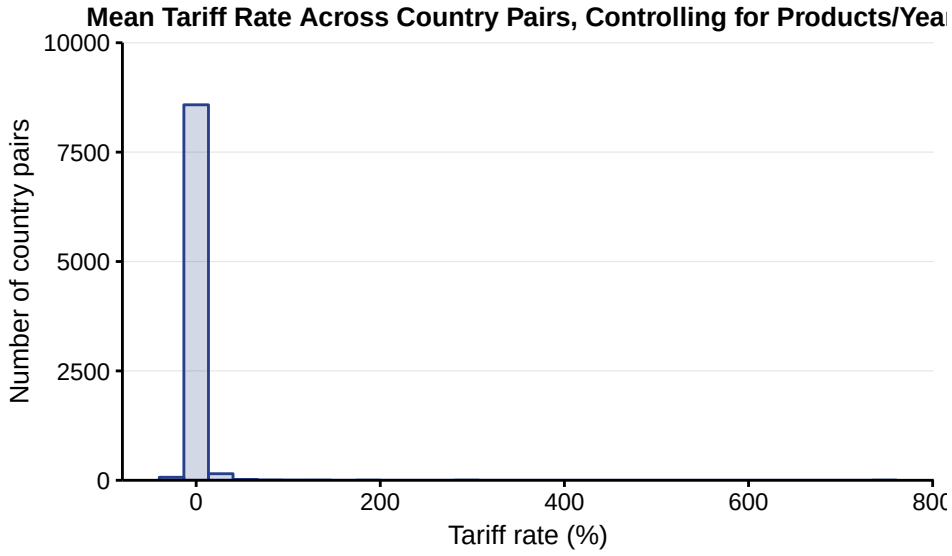
```r
hs12_all_tariffs$Year <- floor_date(hs12_all_tariffs$Year,"year")

hs12_all_tariffs$Year.f <- factor(hs12_all_tariffs$Year)
hs12_all_tariffs$Products.f <- factor(hs12_all_tariffs$ProductCode)

hs12_all_tariffs$Importer.f <- factor(hs12_all_tariffs$`Reporter_ISO_N`)
hs12_all_tariffs$Exporter.f <- factor(hs12_all_tariffs$`Partner Code`)
hs12_all_tariffs$Pairs.f <- with(hs12_all_tariffs, interaction(Importer.f, Exporter.f))

hs12_all_tariffs <- hs12_all_tariffs[, .(Year, SimpleAverage, Year.f, Products.f, Pairs.f)]

reg <- felm(SimpleAverage ~ 1 | Year.f + Products.f + Pairs.f,
            data = hs12_all_tariffs,
            exactDOF = FALSE,
            keepX = FALSE,
            keepCX = FALSE)

fes <- getfe(reg,
             se=TRUE,
             bN = 50
)

Yearfes <- subset(fes,fe == "Year.f")

Yearfes$ci_ub <- Yearfes$effect + (1.96 * Yearfes$se)
Yearfes$ci_lb <- Yearfes$effect - (1.96 * Yearfes$se)
Yearfes <- merge(Yearfes,unique(hs12_all_tariffs[,list(Year,Year.f)]),by.x = "idx",by.y="Year.f")

ggplot(data = Yearfes, aes(Year,effect)) +
  geom_errorbar(aes(ymin = ci_lb, ymax = ci_ub), color = "grey35") +
  geom_line(color = "royalblue4", size = 1) +
  geom_point(color = "royalblue4") +
  background_grid(major = 'y', minor = "none") +
  scale_y_continuous(expand = c(0, 0), limits = c(-3,1), minor_breaks = NULL) +
  xlab(label = "") +
  ylab(label = "Tariff rate (%)") +
  labs(title = "Tariff Rate Over Time, Controlling for Product/Country Pairs")
```



```
productfes <- subset(fes,fe == "Products.f")
productfes <- productfes[,c("effect","idx")]

ggplot(data=productfes, aes(effect)) +
  geom_histogram(col="royalblue4",
                 fill="royalblue4",
                 alpha=.2) +
  background_grid(major = 'y', minor = "none") +
  scale_y_continuous(expand = c(0, 0), limits = c(0,3000), minor_breaks = NULL) +
  labs(title="Mean Tariff Rate Across Products, Controlling for Country Pairs/Years") +
  labs(x="Tariff rate (%)", y="Number of products")
```




```
pairfes <- subset(fes,fe == "Pairs.f")
pairfes <- pairfes[,c("effect","idx")]

ggplot(data=pairfes, aes(effect)) +
  geom_histogram(col="royalblue4",
                 fill="royalblue4",
                 alpha=.2) +
  background_grid(major = 'y', minor = "none") +
  scale_y_continuous(expand = c(0, 0), limits = c(0, 10000), minor_breaks = NULL) +
  labs(title="Mean Tariff Rate Across Country Pairs, Controlling for Products/Years") +
  labs(x="Tariff rate (%)", y="Number of country pairs")
```

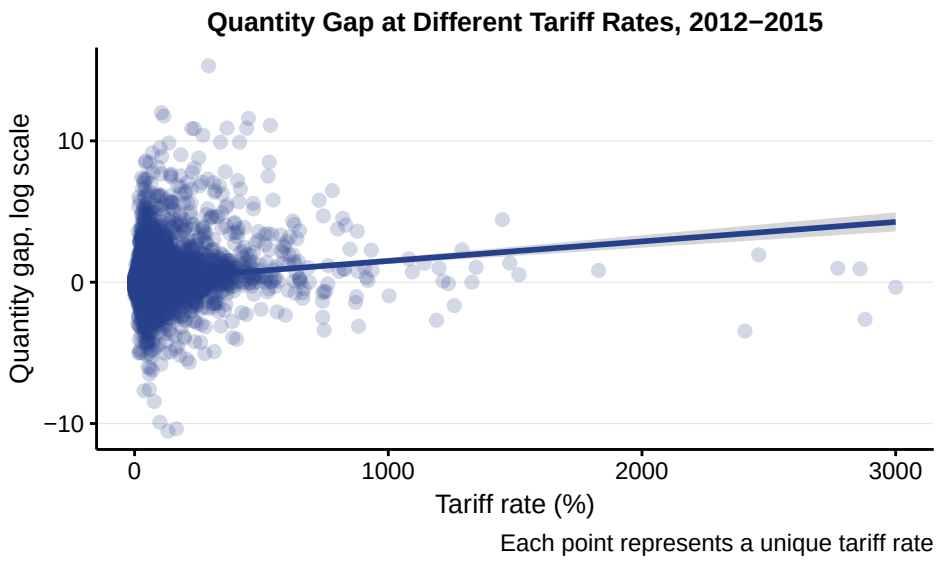



```
rm(fes, hs12_all_tariffs, pairfes, Yearfes, productfes, reg)
```

Tariff vs. Trade Data, Quantity Gap

The following figure plots the quantity evasion gap against mean "Simple Average" tariff rates, grouped by each tariff rate that appears in the data.

```
load(paste(DataPath,"Analysis Data/hs12_all_tariffs_qty.Rda", sep = "/"))

hs12_all_tariffs <- hs12_all_tariffs[,
                    .(Year, ProductCode, Importer, Exporter, SimpleAverage, Qty_log_gap)]

hs12_all_tariffs <- hs12_all_tariffs[!is.na(SimpleAverage)]

tariffs <- hs12_all_tariffs[, .(mean = mean(Qty_log_gap)), by = SimpleAverage]

tariffs <- melt(tariffs, id = "SimpleAverage")
tariffs <- rename(tariffs, "Legend" = "variable")

ggplot(tariffs, aes(SimpleAverage, value, colour=Legend)) +
  geom_point(aes(colour = Legend), size = 2, alpha = .2) +
  geom_smooth(method = "lm") +
  scale_colour_manual(values=c("royalblue4")) +
  labs(title="Quantity Gap at Different Tariff Rates, 2012-2015") +
  background_grid(major = 'y', minor = "none") +
  labs(x="Tariff rate (%)", y="Quantity gap, log scale") +
  labs(caption="Each point represents a unique tariff rate") +
  theme(legend.position="none")
```

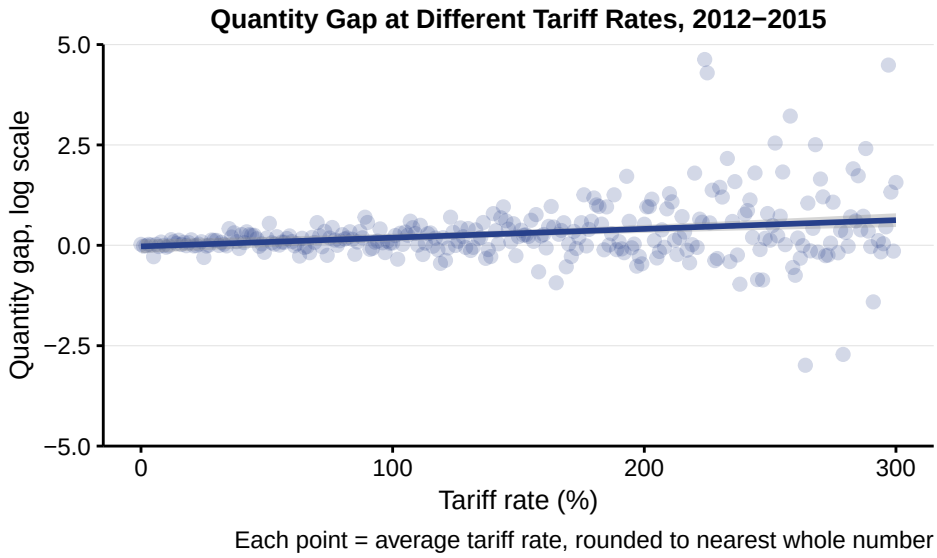


The next figure repeats the first, but with tariff rates grouped to the nearest round number and zoomed in to tariff rates between 0 and 300%.

```
tariffs <- hs12_all_tariffs[SimpleAverage <= 300, ]
tariffs$SimpleAverage <- round(tariffs$SimpleAverage, digits = 0)

tariffs <- tariffs[, .(mean = mean(Qty_log_gap)), by = SimpleAverage]

tariffs <- melt(tariffs, id = "SimpleAverage")
tariffs <- rename(tariffs, "Legend" = "variable")

ggplot(tariffs, aes(SimpleAverage, value, colour=Legend)) +
  geom_point(aes(colour = Legend), size = 2, alpha = .2) +
  geom_smooth(method = "lm") +
  scale_colour_manual(values=c("royalblue4")) +
  labs(title="Quantity Gap at Different Tariff Rates, 2012-2015") +
  scale_y_continuous(expand = c(0, 0), limits = c(-5, 5)) +
  background_grid(major = 'y', minor = "none") +
  labs(x="Tariff rate (%)", y="Quantity gap, log scale") +
  labs(caption="Each point = average tariff rate, rounded to nearest whole number") +
  theme(legend.position="none")
```

```
## Warning: Removed 2 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 2 rows containing missing values (geom_point).
```



Next, a simple regression of the quantity evasion gap regressed on the simple average tariff rate:

```
simpreg = lm(Qty_log_gap ~ SimpleAverage, data = hs12_all_tariffs)
summary(simpreg)
```

```
##
## Call:
## lm(formula = Qty_log_gap ~ SimpleAverage, data = hs12_all_tariffs)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -21.3690  -1.0005   0.0058   1.0323  19.5888
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -5.822e-03  7.464e-04    -7.8 6.19e-15 ***
## SimpleAverage  6.308e-04  5.898e-05    10.7  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.425 on 11401271 degrees of freedom
## Multiple R-squared:  1.003e-05,  Adjusted R-squared:  9.944e-06
## F-statistic: 114.4 on 1 and 11401271 DF,  p-value: < 2.2e-16
```