Summary of Trade and Tariff Data, Value 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 Value 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 value from reported export value.

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

load(paste(DataPath,"Analysis Data/hs12_value.Rda", sep = "/"))
hs12_value <- hs12_value[, .(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_value[, uniqueN(`Commodity Code`), by=Period]
product_year_trade <- rename(product_year_trade, Products_trade = V1)

pair_year_trade <- unique(setDT(hs12_value), 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	5199	4667	6523	7684	0.4841
2013	5200	6123	6524	10262	0.4756
2014	5197	6754	6522	11721	0.4592
2015	5197	7395	6522	12741	0.4625
2016	0	0	6521	6741	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 in tariff data / trade 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_value[, uniqueN(`Period`), by=`Commodity Code`]
year_product_trade <- rename(year_product_trade, Years_trade = V1)

pair_product_trade <- unique(setDT(hs12_value), 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
01	0	0	5	3574	0
0101	0	0	5	1063	0
0102	0	0	5	722	0
0103	0	0	5	422	0
0104	0	0	5	396	0
0105	0	0	5	869	0
0106	0	0	5	2701	0
02	0	0	5	3312	0
0201	0	0	5	1178	0
0202	0	0	5	1515	0

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

ProductCode	Years_tariffs	Pairs_tariffs	Years_trade	Pairs_trade	Coverage
030195	3	3	3	3	1
811213	3	8	3	8	1
811252	2	2	2	2	1
030283	4	14	5	14	0.8
121140	4	13	5	13	0.8
261210	4	9	5	9	0.8
290374	4	8	5	8	0.8
290551	4	5	5	5	0.8
382473	4	10	5	10	0.8
030231	4	145	5	150	0.7733

```
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 in tariff data / trade 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_value[, uniqueN(`Commodity Code`), by = c("Importer", "Exporter")]
product_pair_trade <- rename(product_pair_trade, Products_trade = V1)

year_pair_trade <- hs12_value[, 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:        Austria        Germany             4783             4
##  2:        Belgium    Netherlands             4779             4
##  3:    Netherlands        Germany             4726             4
##  4:        Germany    Netherlands             4688             4
##  5:    Netherlands        Belgium             4678             4
##  6:         France    Netherlands             4570             4
##  7:          Italy    Netherlands             4487             4
##  8: United Kingdom    Netherlands             4410             4
```

```
##  9:    Netherlands United Kingdom             4378             4
## 10:          Spain    Netherlands             4319             4
##     Products_trade Years_trade  Coverage
##  1:           6071           4 0.7878438
##  2:           6071           4 0.7871850
##  3:           6009           4 0.7864869
##  4:           5967           4 0.7856544
##  5:           5960           4 0.7848993
##  6:           5844           4 0.7819986
##  7:           5750           4 0.7803478
##  8:           5666           4 0.7783269
##  9:           5633           4 0.7772058
## 10:           5560           4 0.7767986
```

```
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_value[, .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	254652	377941	0.6738
Netherlands	329974	490354	0.6729
Finland	200838	301229	0.6667
Slovenia	179426	270490	0.6633
United Arab Emirates	262670	398288	0.6595
Ukraine	81721	126937	0.6438
Kuwait	102250	163715	0.6246
Bangladesh	61777	100614	0.614
Qatar	98262	160702	0.6115
Papua New Guinea	10667	17565	0.6073

No country has tariffs for more than 70% 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_value[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
```

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

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

```
##                            Importer Tariffs  Trade Share_covered
##   1:                       Australia  241533 241533     1.0000000
##   2:                         Austria  254652 254652     1.0000000
##   3:                         Bahrain   72995  72995     1.0000000
##   4:                      Bangladesh   61777  61777     1.0000000
##   5:                         Belarus  115354 115354     1.0000000
##   6:                         Belgium  292977 292977     1.0000000
##   7:                         Bermuda   15800  15800     1.0000000
##   8:              Bosnia Herzegovina  116558 116558     1.0000000
##   9:                        Botswana   23502  23502     1.0000000
##  10:                        Bulgaria  182509 182509     1.0000000
##  11:                         Burundi    5545   5545     1.0000000
##  12:                          Canada  260049 260049     1.0000000
##  13:                         Croatia  178093 178093     1.0000000
##  14:                          Cyprus  100172 100172     1.0000000
##  15:                         Czechia  246562 246562     1.0000000
##  16:                         Denmark  230043 230043     1.0000000
##  17:                           Egypt   65823  65823     1.0000000
##  18:                         Estonia  156658 156658     1.0000000
##  19:                         Finland  200838 200838     1.0000000
##  20:                          France  387093 387093     1.0000000
##  21:                         Germany  461769 461769     1.0000000
##  22:                          Greece  187390 187390     1.0000000
##  23:                         Hungary  208717 208717     1.0000000
##  24:                         Iceland   87525  87525     1.0000000
##  25:                         Ireland  145086 145086     1.0000000
##  26:                          Israel  165226 165226     1.0000000
##  27:                           Italy  359367 359367     1.0000000
##  28:                      Kazakhstan  136323 136323     1.0000000
##  29:                          Kuwait  102250 102250     1.0000000
##  30:                      Kyrgyzstan   12641  12641     1.0000000
##  31:                          Latvia  165669 165669     1.0000000
##  32:                       Lithuania  188678 188678     1.0000000
##  33:                      Luxembourg  108464 108464     1.0000000
##  34:                           Malta   82210  82210     1.0000000
##  35:                       Mauritius   77813  77813     1.0000000
##  36:                      Montenegro   68749  68749     1.0000000
##  37:                         Namibia   41295  41295     1.0000000
##  38:                           Nepal   25808  25808     1.0000000
##  39:                     Netherlands  329974 329974     1.0000000
##  40:                     New Zealand  165168 165168     1.0000000
##  41:                          Norway  222210 222210     1.0000000
##  42:                        Pakistan  101729 101729     1.0000000
##  43:                Papua New Guinea   10667  10667     1.0000000
##  44:                          Poland  273666 273666     1.0000000
##  45:                        Portugal  180956 180956     1.0000000
##  46:                           Qatar   98262  98262     1.0000000
##  47:                  Rep. of Korea  230299 230299     1.0000000
##  48:                 Rep. of Moldova   21631  21631     1.0000000
##  49:                         Romania  239718 239718     1.0000000
##  50:              Russian Federation  273722 273722     1.0000000
##  51:                          Rwanda   20016  20016     1.0000000
```

```
##  52:                           Samoa   13337  13337     1.0000000
##  53:                       Singapore  246924 246924     1.0000000
##  54:                        Slovakia  173675 173675     1.0000000
##  55:                        Slovenia  179426 179426     1.0000000
##  56:                    Solomon Isds    3497   3497     1.0000000
##  57:                    South Africa  218019 218019     1.0000000
##  58:                           Spain  307861 307861     1.0000000
##  59:              State of Palestine    3483   3483     1.0000000
##  60:                          Sweden  244356 244356     1.0000000
##  61:                     Switzerland  294810 294810     1.0000000
##  62:              TFYR of Macedonia   99387  99387     1.0000000
##  63:                             USA  395951 395951     1.0000000
##  64:                          Uganda   46524  46524     1.0000000
##  65:                         Ukraine   81721  81721     1.0000000
##  66:           United Arab Emirates  262670 262670     1.0000000
##  67:                 United Kingdom  352156 352156     1.0000000
##  68:        United Rep. of Tanzania   59948  59948     1.0000000
##  69:                        Thailand  166277 213135     0.7801487
##  70:                      Azerbaijan   52536  70321     0.7470884
##  71:                           Japan  162328 217993     0.7446478
##  72:                    Saudi Arabia   87091 118643     0.7340593
##  73:                         Ecuador   52947  80130     0.6607638
##  74:                           Palau    1831   3269     0.5601101
##  75:                           China  154108 294826     0.5227083
##  76:                         Georgia   52890 102899     0.5139992
##  77:                          Jordan   38887  78830     0.4933020
##  78:                            Fiji   17624  38094     0.4626450
##  79:                           Aruba    2599   6661     0.3901817
##  80:                         Lebanon   29015  83002     0.3495699
##  81:                         Armenia   13217  39837     0.3317770
##  82:              Brunei Darussalam   11353  34561     0.3284917
##  83:                           India   57329 174814     0.3279428
##  84:                       Sri Lanka   24406  74724     0.3266153
##  85:                        Malaysia   48522 148582     0.3265671
##  86:                        Zimbabwe   10518  32936     0.3193466
##  87:                            Oman   24502  91471     0.2678663
##  88:                          Turkey   57797 222416     0.2598599
##  89:                         Albania       0  50837     0.0000000
##  90:                         Algeria       0  98410     0.0000000
##  91:                         Andorra       0  19947     0.0000000
##  92:                          Angola       0  58484     0.0000000
##  93:                       Argentina       0 116742     0.0000000
##  94:                         Bahamas       0  14231     0.0000000
##  95:                          Belize       0  10342     0.0000000
##  96:                           Benin       0   5132     0.0000000
##  97:                          Bhutan       0    863     0.0000000
##  98: Bolivia (Plurinational State of)       0  76580     0.0000000
##  99:                          Brazil       0 194706     0.0000000
## 100:                    Burkina Faso       0   6232     0.0000000
## 101:                      Cabo Verde       0  17642     0.0000000
## 102:                        Cambodia       0  36409     0.0000000
## 103:                        Cameroon       0  43275     0.0000000
## 104:                           Chile       0 170307     0.0000000
## 105:            China, Hong Kong SAR       0 178908     0.0000000
## 106:                China, Macao SAR       0  19767     0.0000000
## 107:                        Colombia       0 143605     0.0000000
## 108:                           Congo       0  11905     0.0000000
```

```
## 109:                      Costa Rica       0 101660     0.0000000
## 110:                   CÃ´te d'Ivoire       0  17758     0.0000000
## 111:                   Dominican Rep.       0  84338     0.0000000
## 112:                      El Salvador       0  75058     0.0000000
## 113:                         Ethiopia       0  29026     0.0000000
## 114:                 French Polynesia       0  37707     0.0000000
## 115:                            Ghana       0  23488     0.0000000
## 116:                        Greenland       0  14503     0.0000000
## 117:                        Guatemala       0  96723     0.0000000
## 118:                           Guinea       0  19138     0.0000000
## 119:                         Honduras       0  32169     0.0000000
## 120:                        Indonesia       0 169924     0.0000000
## 121:                          Jamaica       0  19111     0.0000000
## 122:                       Madagascar       0  42947     0.0000000
## 123:                           Malawi       0  14769     0.0000000
## 124:                         Maldives       0  44450     0.0000000
## 125:                           Mexico       0 205032     0.0000000
## 126:                         Mongolia       0  37371     0.0000000
## 127:                          Morocco       0  33630     0.0000000
## 128:                    New Caledonia       0  22369     0.0000000
## 129:                        Nicaragua       0  56752     0.0000000
## 130:                            Niger       0   4317     0.0000000
## 131:                  Other Asia, nes       0 139999     0.0000000
## 132:                           Panama       0  87073     0.0000000
## 133:                         Paraguay       0  66969     0.0000000
## 134:                             Peru       0 132629     0.0000000
## 135:                          Senegal       0  14216     0.0000000
## 136:                           Serbia       0 158266     0.0000000
## 137:                            Sudan       0   6889     0.0000000
## 138:                             Togo       0   4978     0.0000000
## 139:                            Tonga       0   2986     0.0000000
## 140:                          Tunisia       0  99801     0.0000000
## 141:                          Uruguay       0  95342     0.0000000
## 142:                         Viet Nam       0 153291     0.0000000
## 143:                            Yemen       0  24754     0.0000000
## 144:                           Zambia       0  50618     0.0000000
##                              Importer Tariffs  Trade Share_covered
```

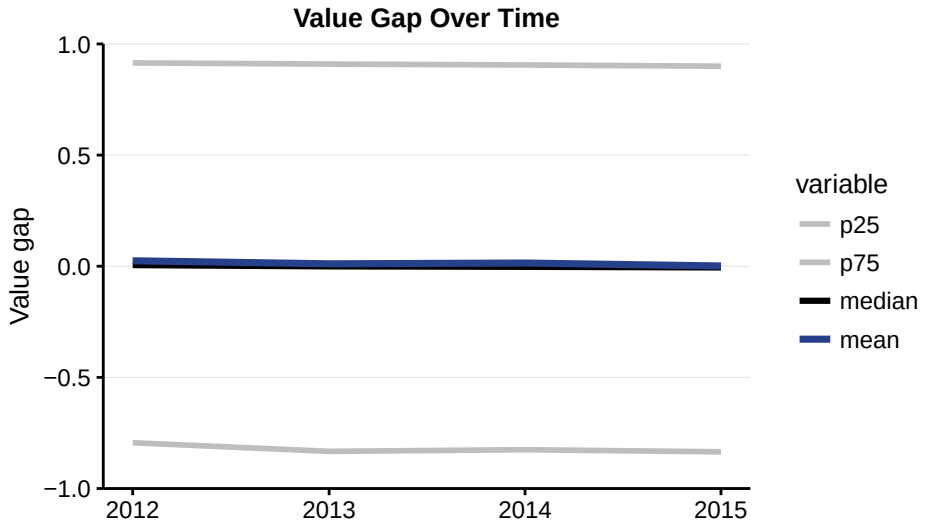
```
rm(hs12_all_tariffs, hs12_value, matches, tariffs, trade)
```

Value Trade Gap, Limited to Data with Corresponding Tariff Rates

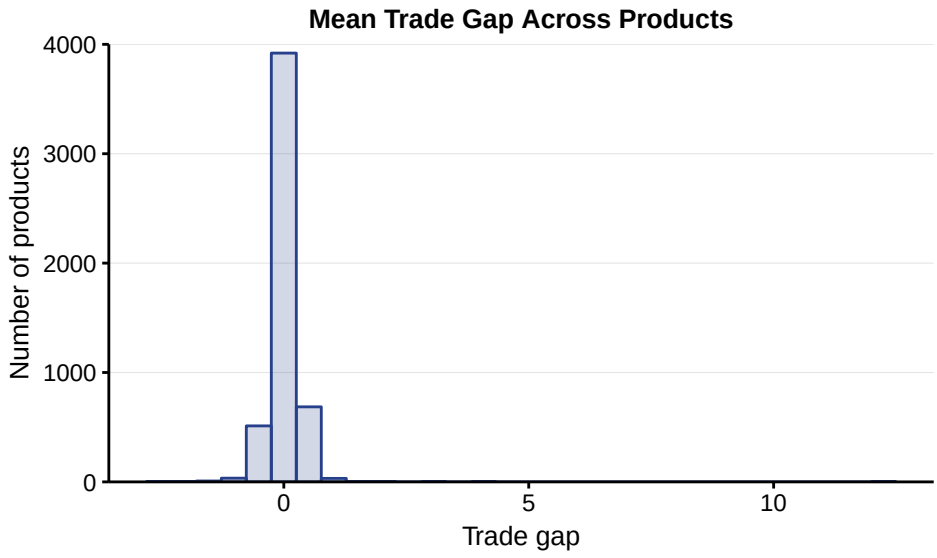
The following figures are the same as the previous summary files on the evasion gap in U.S. dollars (log exports - log imports), but limited to data with tariff rates.

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

load(paste(DataPath,"Analysis Data/hs12_all_tariffs.Rda", sep = "/"))

hs12_all_tariffs <- hs12_all_tariffs[, .(Year, ProductCode, Importer, Exporter, 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(Log_gap)),
                          median = as.double(median(Log_gap)),
```

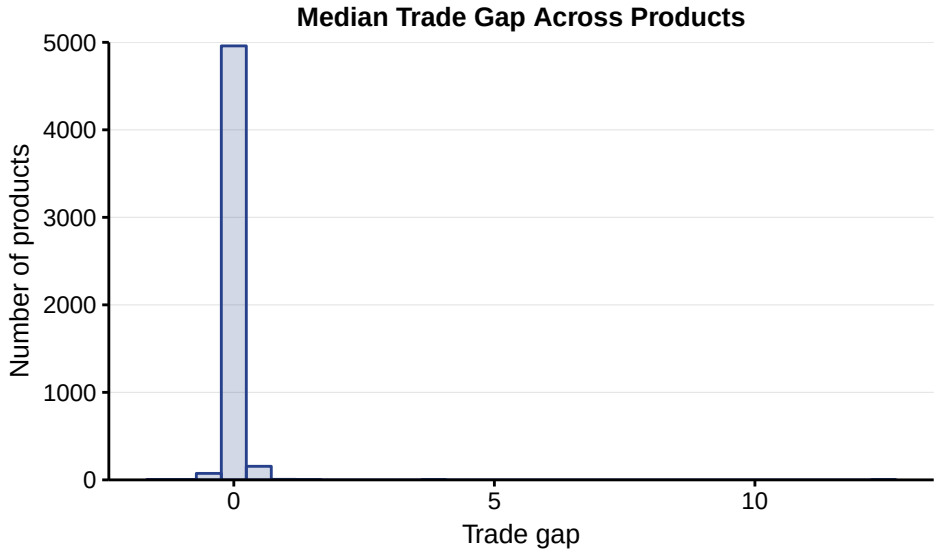
```
                          p25 = as.double(quantile(Log_gap,.25)),
                          p75 = as.double(quantile(Log_gap,.75))
),
by=Year]

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(-1,1), minor_breaks = NULL) +
  xlab(label = "") +
  ylab(label = "Value gap") +
  labs(title="Value Gap Over Time")
```



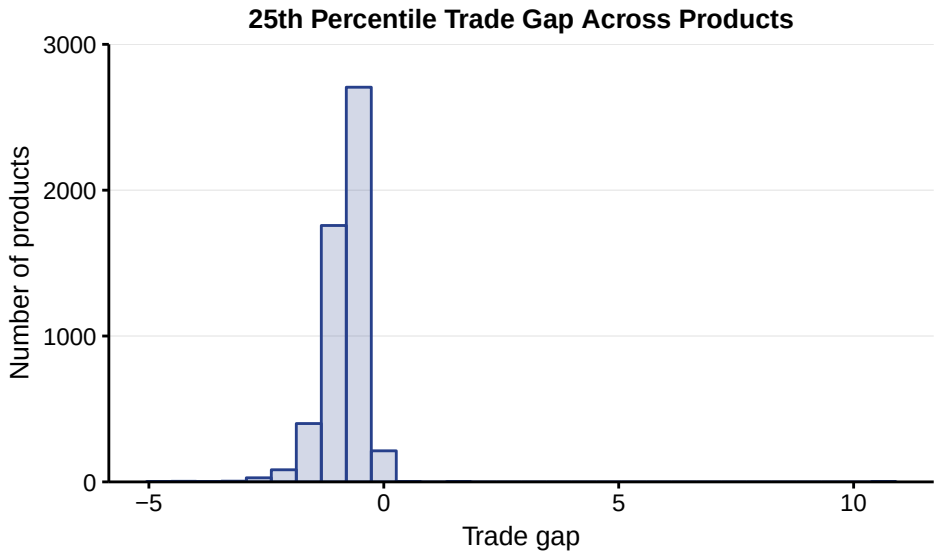

```
#Across products?
products <- hs12_all_tariffs[, .(mean = as.double(mean(Log_gap)),
                          median = as.double(median(Log_gap)),
                          p25 = as.double(quantile(Log_gap,.25)),
                          p75 = as.double(quantile(Log_gap,.75))
),
by= ProductCode]

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, 4000))  +
  labs(title="Mean Trade Gap Across Products") +
  labs(x="Trade gap", y="Number of products")
```



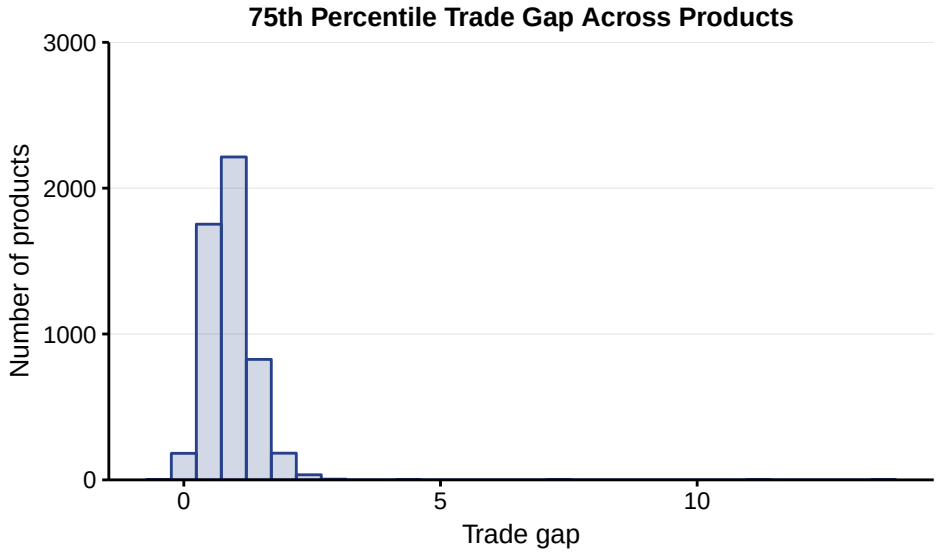

```
ggplot(data=products, aes(median)) +
  geom_histogram(col="royalblue4",
                 fill="royalblue4",
                 alpha=.2) +
  background_grid(major = 'y', minor = "none") +
  scale_y_continuous(expand = c(0, 0), limits = c(0, 5000)) +
  labs(title="Median Trade Gap Across Products") +
  labs(x="Trade 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(0,3000),  minor_breaks = NULL) +
  labs(title="25th Percentile Trade Gap Across Products") +
  labs(x="Trade gap", y="Number of products")
```

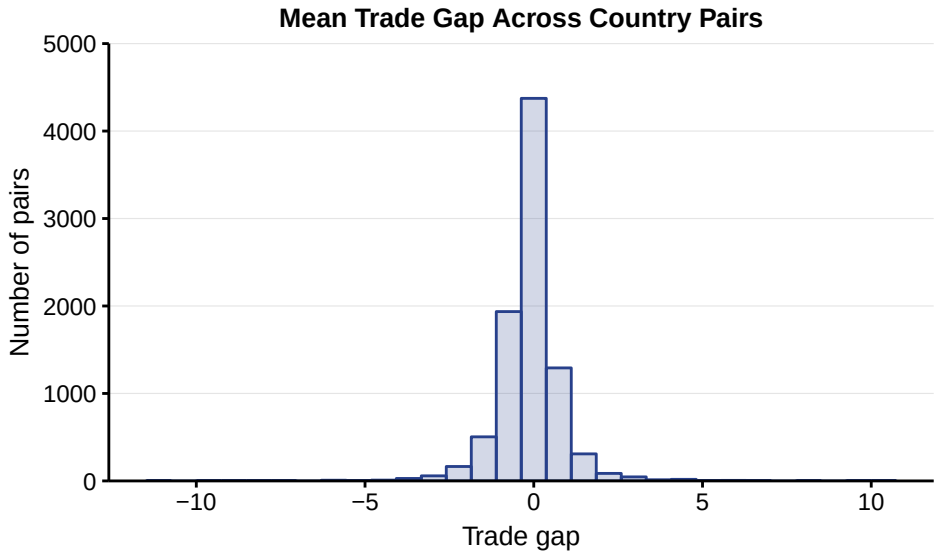



```
ggplot(data=products, aes(p75)) +
  geom_histogram(col="royalblue4",
                 fill="royalblue4",
                 alpha=.2) +
  background_grid(major = 'y', minor = "none") +
  scale_y_continuous(expand = c(0, 0),  limits = c(0, 3000), minor_breaks = NULL) +
  labs(title="75th Percentile Trade Gap Across Products") +
  labs(x="Trade gap", y="Number of products")
```

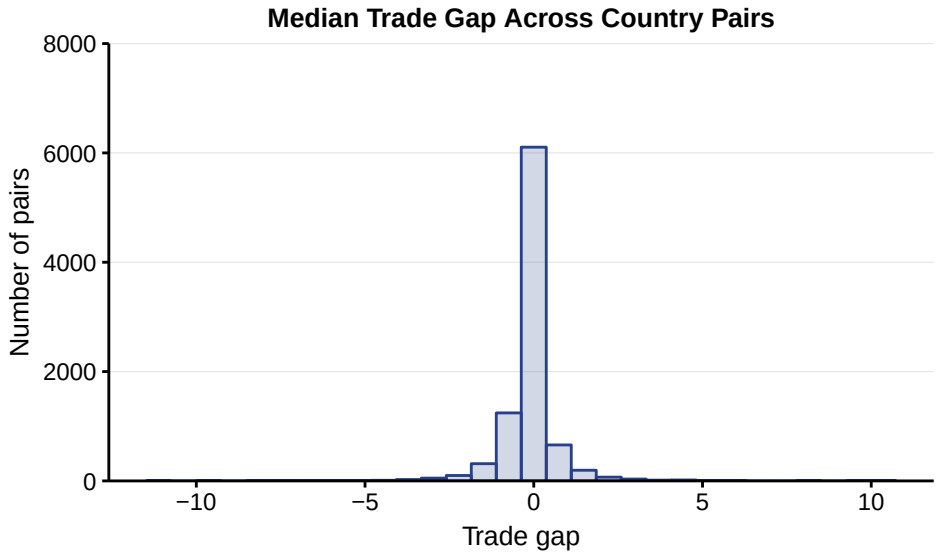



```
#Across countries?
countries <- hs12_all_tariffs[, .(mean = as.double(mean(Log_gap)),
                            median = as.double(median(Log_gap)),
                            p25 = as.double(quantile(Log_gap,.25)),
                            p75 = as.double(quantile(Log_gap,.75))
),
by= c("Importer", "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 Trade Gap Across Country Pairs") +
  labs(x="Trade gap", y="Number of pairs")
```

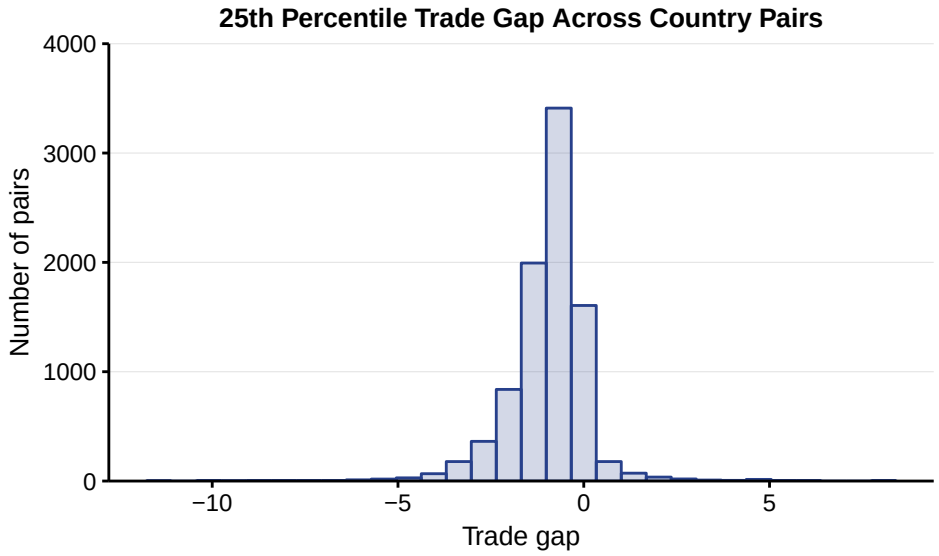



```
ggplot(data=countries, aes(median)) +
  geom_histogram(col="royalblue4",
                 fill="royalblue4",
                 alpha=.2) +
  background_grid(major = 'y', minor = "none") +
  scale_y_continuous(expand = c(0, 0), limits = c(0, 8000), minor_breaks = NULL) +
  labs(title="Median Trade Gap Across Country Pairs") +
  labs(x="Trade gap", y="Number of pairs")
```

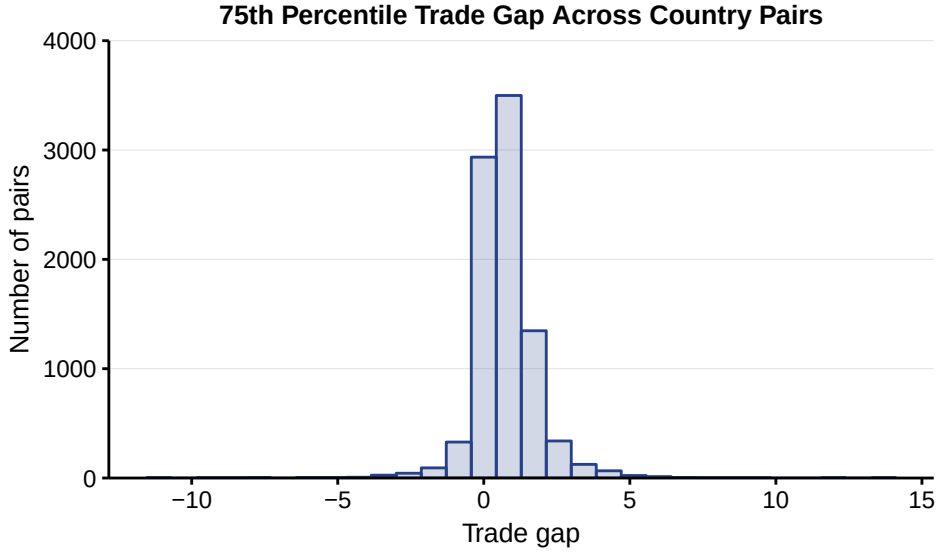



```
ggplot(data=countries, aes(p25)) +
  geom_histogram(col="royalblue4",
                 fill="royalblue4",
                 alpha=.2) +
```

```
  background_grid(major = 'y', minor = "none") +
  scale_y_continuous(expand = c(0, 0), limits = c(0, 4000),  minor_breaks = NULL) +
  labs(title="25th Percentile Trade Gap Across Country Pairs") +
  labs(x="Trade 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 Trade Gap Across Country Pairs") +
  labs(x="Trade 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.Rda", sep = "/"))
```

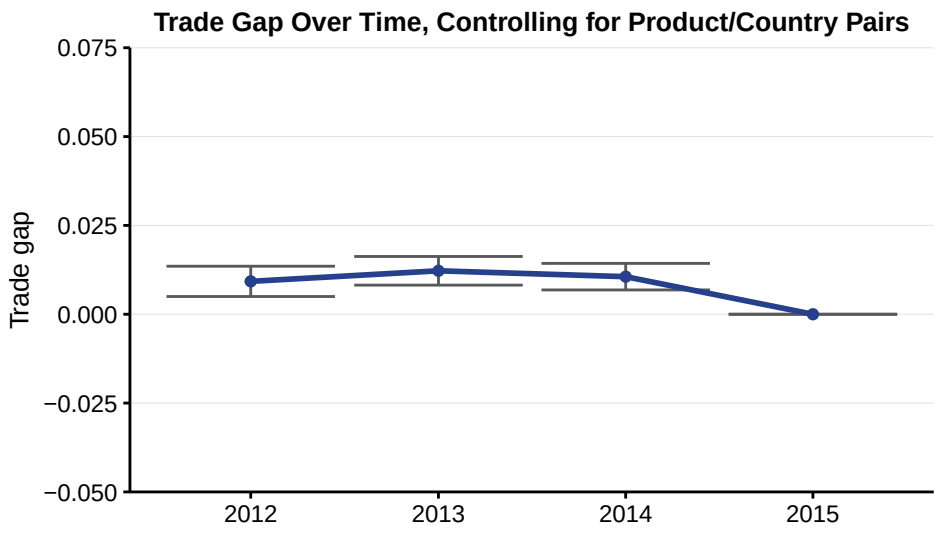
```r
hs12_all_tariffs <- hs12_all_tariffs[, .(Year, ProductCode, `Reporter_ISO_N`, `Partner Code`, Log_gap)]

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, Log_gap, Year.f, Products.f, Pairs.f)]

reg <- felm(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
)
```

```
## Warning in is.estimable(ef, obj$fe): non-estimable function, largest error
## 5e-06 in coordinate 7179 ("Pairs.f.792.275")
```

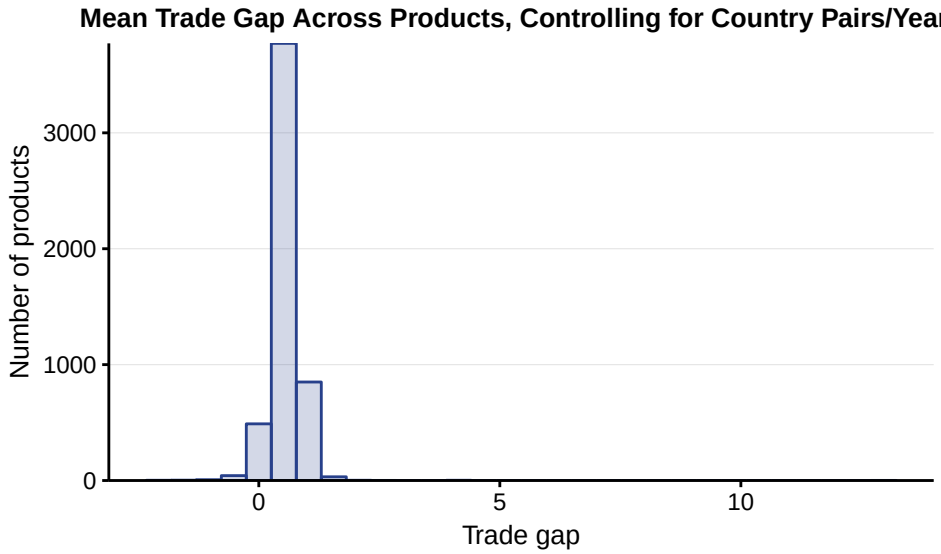
```
## Warning in getfe.kaczmarz(obj, se, ef = ef, bN = bN, robust = robust,
## cluster = cluster, : Supplied function seems non-estimable
```

```
## ...finished 56 of 56 vectors in 317 seconds
```

```r
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(-.050,.075), minor_breaks = NULL) +
  xlab(label = "") +
  ylab(label = "Trade gap") +
  labs(title = "Trade Gap Over Time, Controlling for Product/Country Pairs")
```

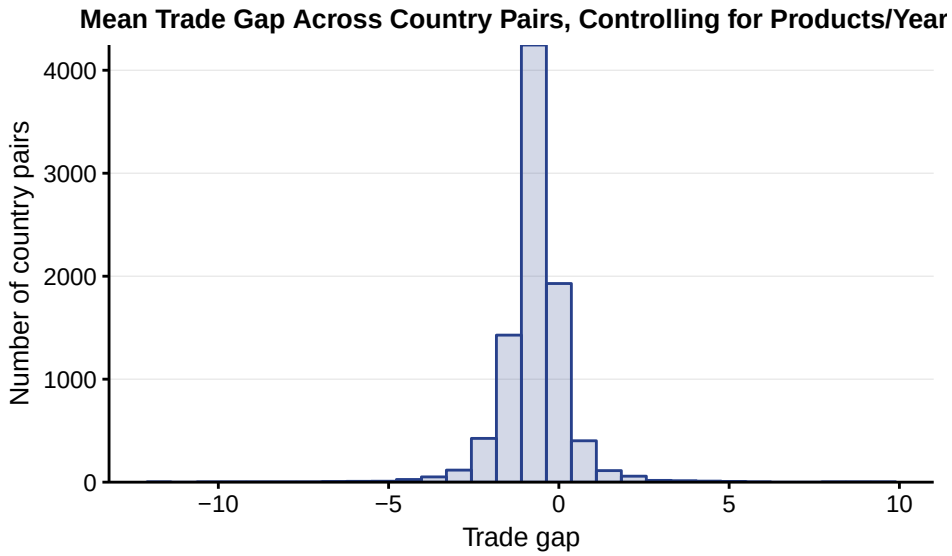



```
productfes <- subset(fes,fe == "Products.f")
productfes <- productfes[,c("effect","idx")]

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), minor_breaks = NULL) +
  labs(title="Mean Trade Gap Across Products, Controlling for Country Pairs/Years") +
  labs(x="Trade 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), minor_breaks = NULL) +
  labs(title="Mean Trade Gap Across Country Pairs, Controlling for Products/Years") +
  labs(x="Trade gap", y="Number of country pairs")
```



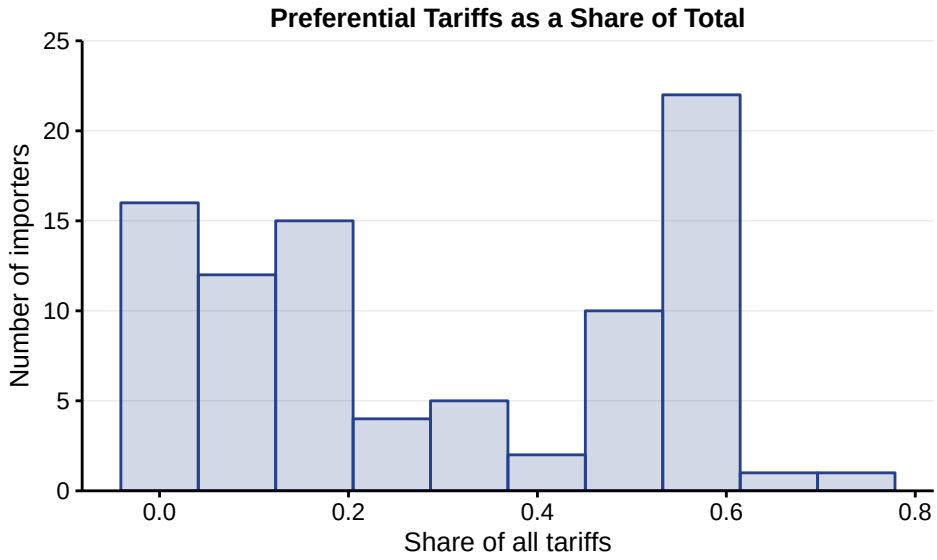

```
rm(fes, hs12_all_tariffs, pairfes, Yearfes, productfes, reg)
```

Tariff Data Summary

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

```
load(paste(DataPath,"Analysis Data/hs12_all_tariffs.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") +
```

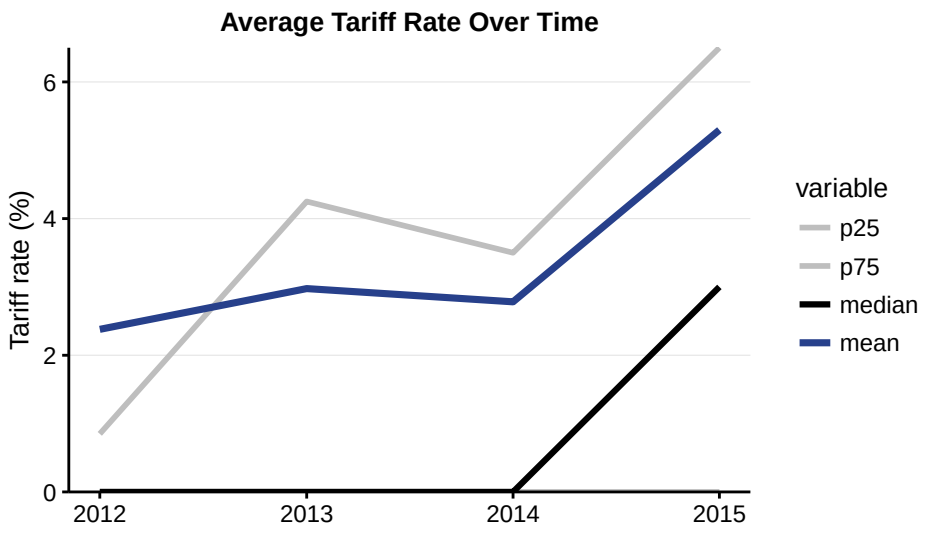
```
  labs(x="Share of all tariffs", y="Number of importers")
```



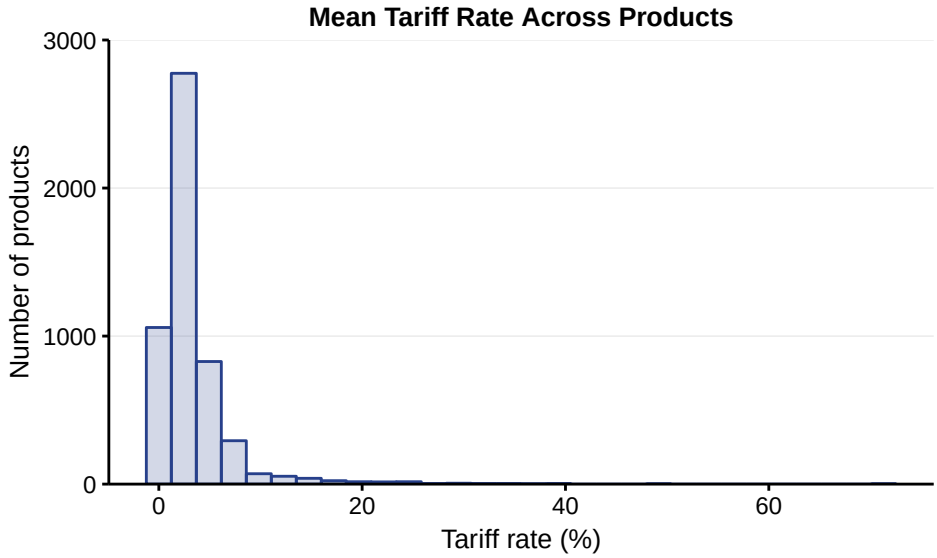

```
rm(hs12_all_tariffs, mfn, pref, tariffs)
```

The next section summarizes the "Simple Average" tariff rate as reported by WITS, for tariffs that match with the value 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.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")
```

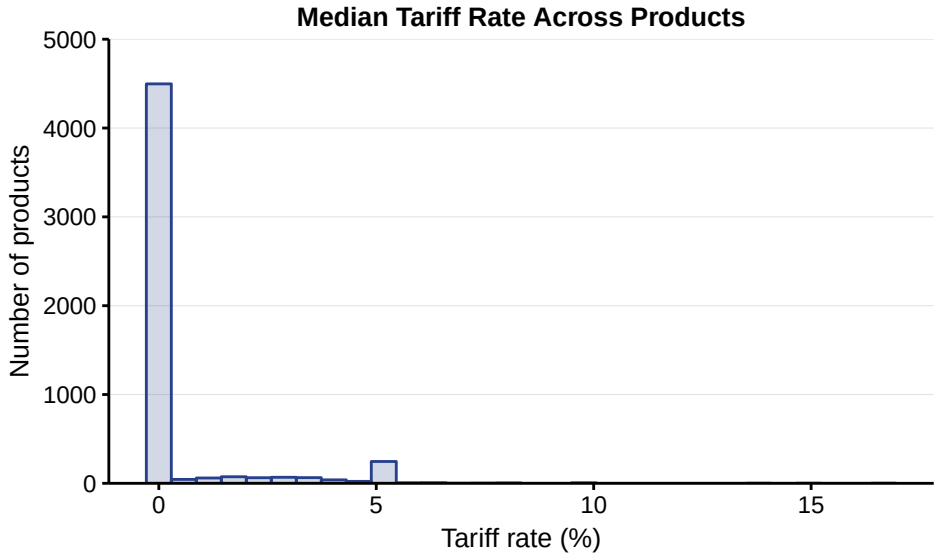



```
#Across products?
products <- hs12_all_tariffs[, .(mean = as.double(mean(SimpleAverage)),
                                 median = as.double(median(SimpleAverage)),
                                 p25 = as.double(quantile(SimpleAverage,.25)),
                                 p75 = as.double(quantile(SimpleAverage,.75))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

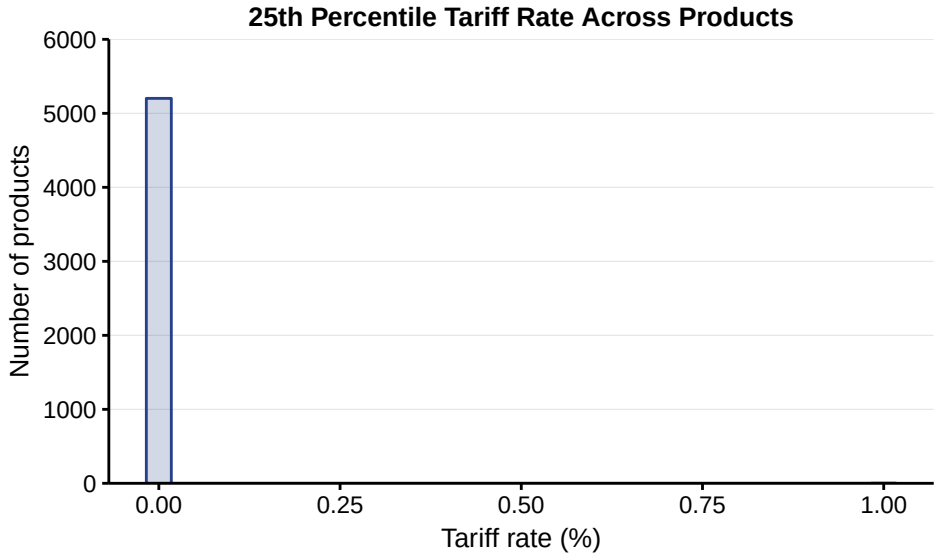



```
ggplot(data=products, aes(median)) +
  geom_histogram(col="royalblue4",
                 fill="royalblue4",
```

```
                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")
```

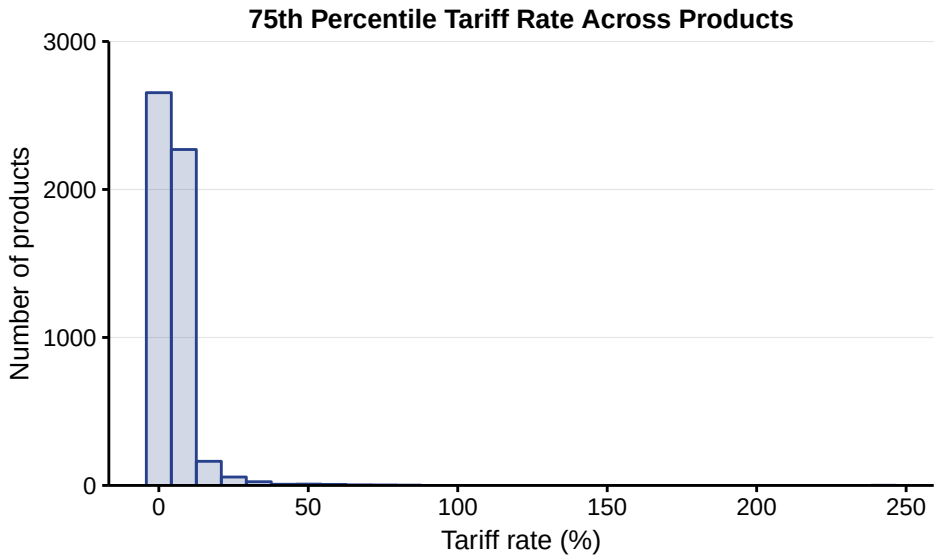



```
ggplot(data=products, aes(p25)) +
  geom_histogram(col="royalblue4",
                 fill="royalblue4",
                 alpha=.2) +
  background_grid(major = 'y', minor = "none") +
  scale_y_continuous(expand = c(0, 0), limits = c(0, 6000), minor_breaks = NULL) +
  labs(title="25th Percentile Tariff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

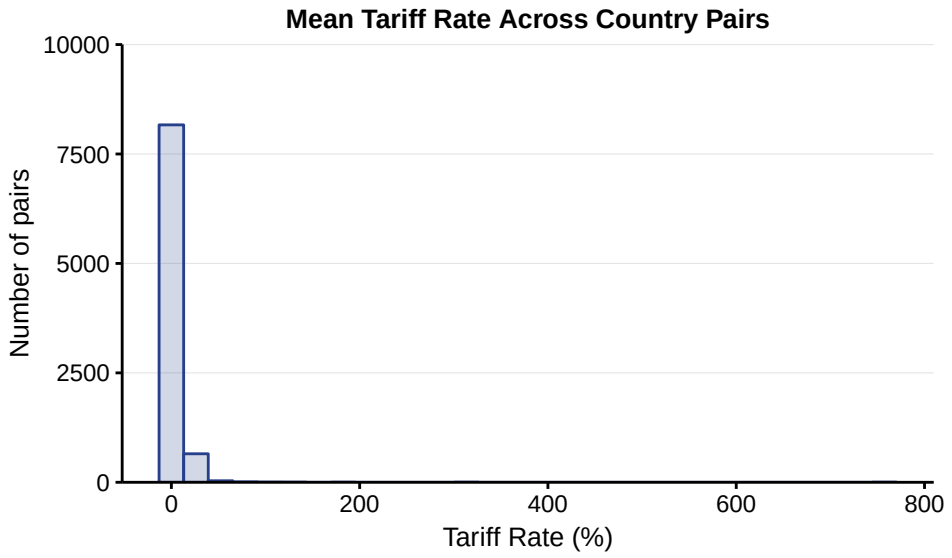



```
ggplot(data=products, aes(p75)) +
  geom_histogram(col="royalblue4",
                 fill="royalblue4",
                 alpha=.2) +
```

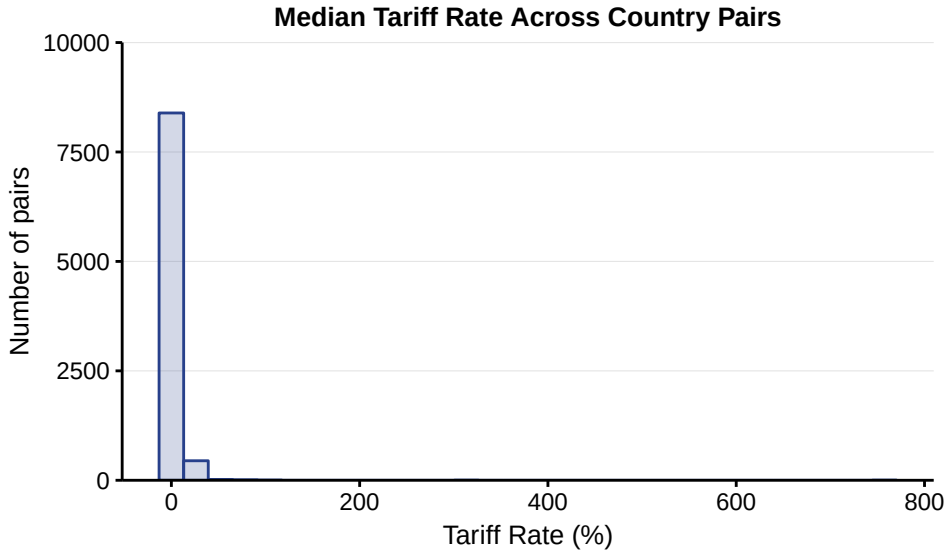
```
  background_grid(major = 'y', minor = "none") +
  scale_y_continuous(expand = c(0, 0), limits = c(0, 3000),  minor_breaks = NULL) +
  labs(title="75th Percentile Tariff Rate Across Products") +
  labs(x="Tariff rate (%)", y="Number of products")
```



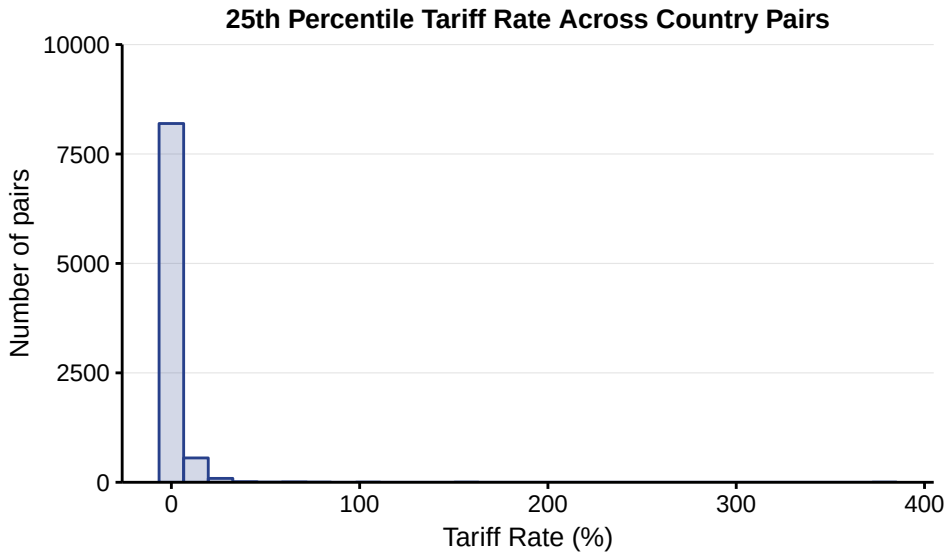

```
#Across countries?
countries <- hs12_all_tariffs[, .(mean = as.double(mean(SimpleAverage)),
                                  median = as.double(median(SimpleAverage)),
                                  p25 = as.double(quantile(SimpleAverage,.25)),
                                  p75 = as.double(quantile(SimpleAverage,.75))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



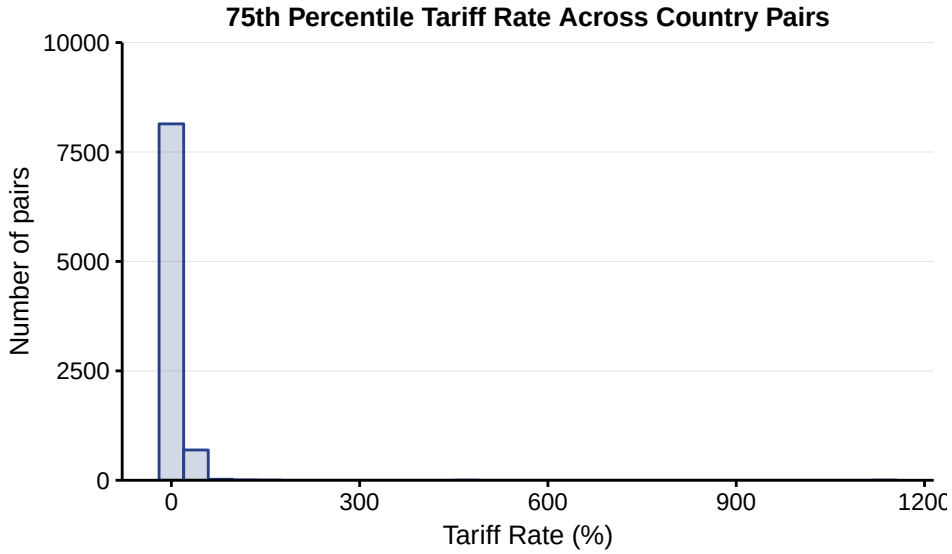

```
ggplot(data=countries, aes(median)) +
  geom_histogram(col="royalblue4",
                 fill="royalblue4",
                 alpha=.2) +
  background_grid(major = 'y', minor = "none") +
  scale_y_continuous(expand = c(0, 0), limits = c(0, 10000), minor_breaks = NULL) +
  labs(title="Median Tariff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

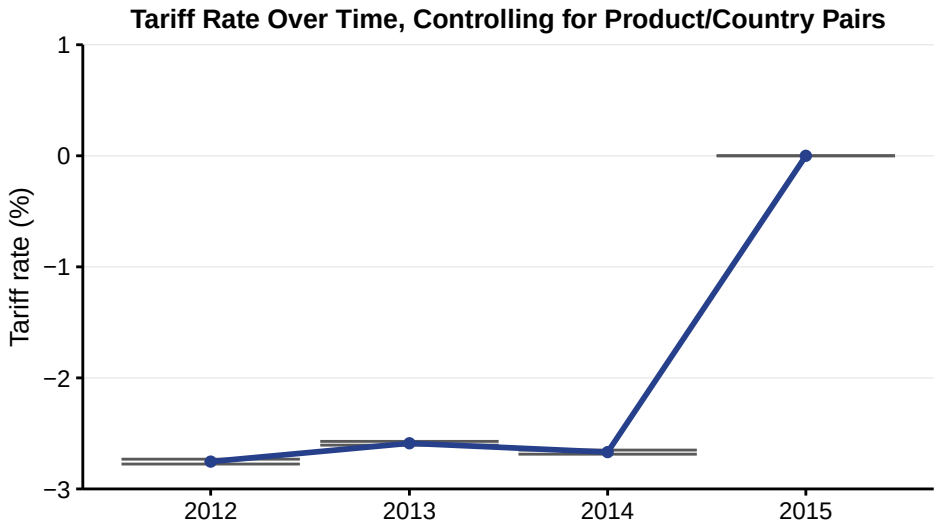



```
rm(Years, products, countries, hs12_all_tariffs)

#Regress trade gap on dummies and plot coefficients

load(paste(DataPath,"Analysis Data/hs12_all_tariffs.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")
```

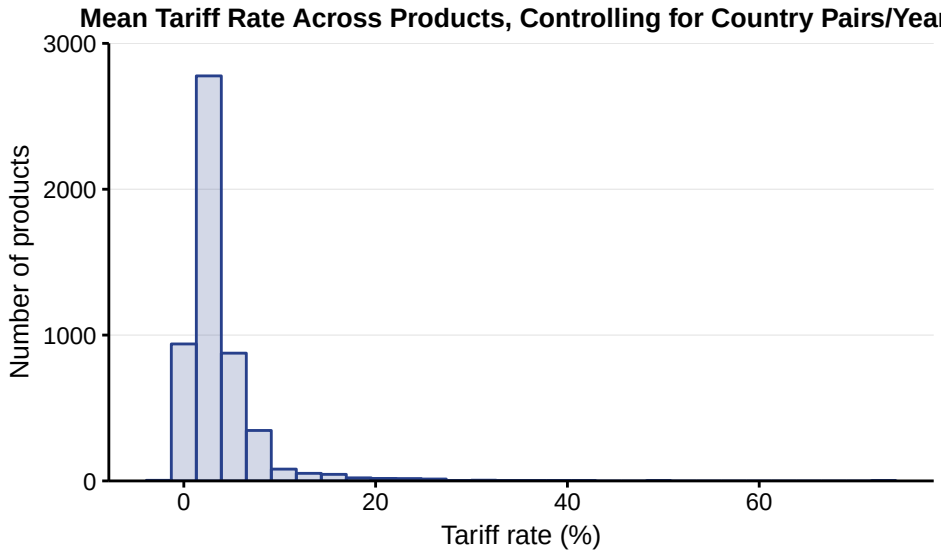
```r
hs12_all_tariffs$Year <- floor_date(hs12_all_tariffs$Year,"year")

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
)
```

```
## ...finished 56 of 56 vectors in 316 seconds
```

```r
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")
```

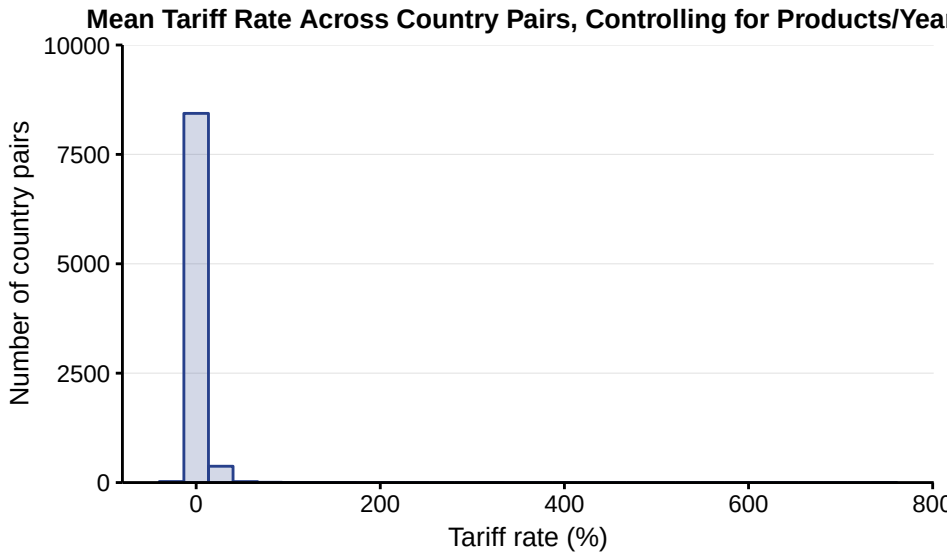



```
productfes <- subset(fes,fe == "Products.f")
productfes <- productfes[,c("effect","idx")]

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")
```

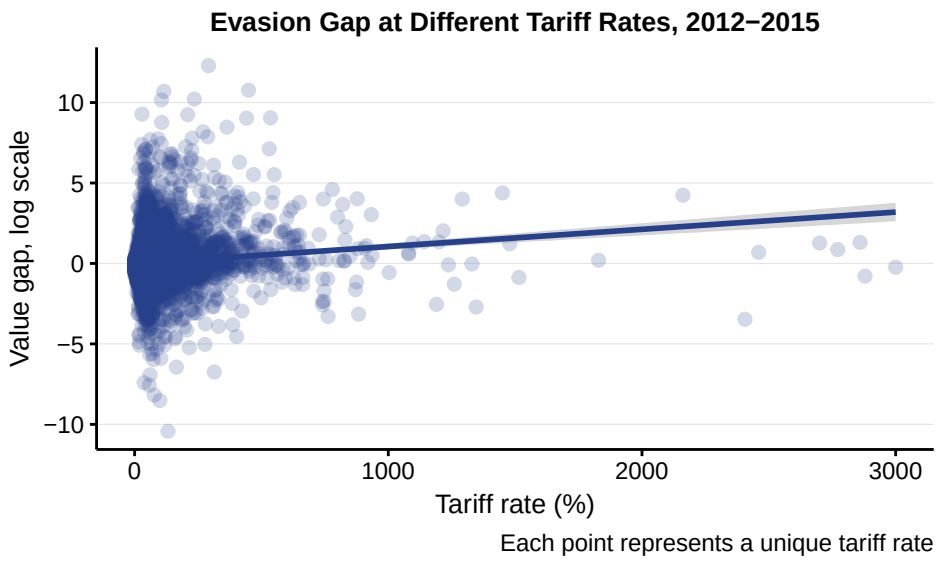



```
rm(fes, hs12_all_tariffs, pairfes, Yearfes, productfes, reg)
```

Tariff vs. Trade Data, Value Gap

The following figure plots the value 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.Rda", sep = "/"))

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

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

tariffs <- hs12_all_tariffs[, .(mean = mean(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="Evasion Gap at Different Tariff Rates, 2012-2015") +
  background_grid(major = 'y', minor = "none") +
  labs(x="Tariff rate (%)", y="Value gap, log scale") +
  labs(caption="Each point represents a unique tariff rate") +
  theme(legend.position="none")
```

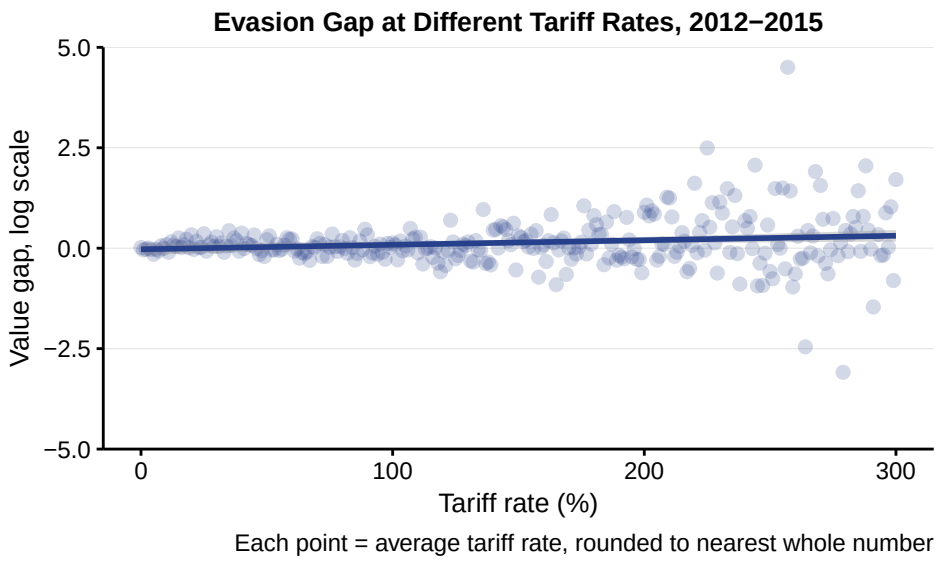


The next figure repeats the first, but with tariff rates grouped to the nearest round number and zoomed in to tariff rates between 0 and 300%.

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

tariffs <- tariffs[, .(mean = mean(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="Evasion 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="Value gap, log scale") +
  labs(caption="Each point = average tariff rate, rounded to nearest whole number") +
  theme(legend.position="none")
```

```
## Warning: Removed 1 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 1 rows containing missing values (geom_point).
```



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

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

```
##
## Call:
## lm(formula = Log_gap ~ SimpleAverage, data = hs12_all_tariffs)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -20.9576  -0.8358  -0.0113   0.8953  17.9454
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   5.317e-03  6.573e-04   8.089 6.01e-16 ***
## SimpleAverage 1.885e-03  5.172e-05  36.447  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.195 on 12034412 degrees of freedom
## Multiple R-squared:  0.0001104,  Adjusted R-squared:  0.0001103
## F-statistic:  1328 on 1 and 12034412 DF,  p-value: < 2.2e-16
```