Recreating the Interconnection Report Plots

I read in the data from M-Lab using the methods that are now available in the Jupiter notebook.

I used the dates 01/01/2013 and 01/01/2014, New York city for M-Lab server location and ASN=174 for Cogent.

The data frame I got from M-Lab has the following variables:

```
 "log_time"          "client_city"          "client_area_code"
"client_ip"          "MLab_ip"              "download_Mbps"
```

I then used https://www.team-cymru.org website to get the ASNs for the `client_ips`.

I joined the two data frames by `client_ips` and got a data frame, lets call it `df_plus`, with these variables:

```
log_time            client_city        client_area_code client_ip
MLab_ip             download_Mbps      client_ASN        client_AS
```

I added a `date` column to the data frame, `df_plus` by converting the `log_time` variable to date, so now the columns of `df_plus` are:

```
"log_time"          "client_city"          "client_area_code"
"client_ip"         "MLab_ip"              "download_Mbps"
"client_ASN"        "client_AS"            "date"
```

Next, I used regular expressions to locate the words "Comcast" in the `client_ASN` variable and filtered `df_plus` for it, creating `comcast_df`.
Then, I filtered `df_plus`'s `client_ASN` variable for "Verizon" and created `verizon_df` and so on, until I had a `time_warner_df` and `cablevision_df` as well.

The next step was to restrict the clients to clients located in the New York City area. I ended up using the `client_area_code` variable, and I filtered `comcast_df`, `verizon_df, time_warner_df` and `cablevision_df` for the following client area codes (chosen by me from the NYC area):

212 332 347 516 631 646 718 845 914 917 929 934

and created `comcast_nyc_area_df, verizon_nyc_area_df,` `time_warner_nyc_area_df,` and `cablevision_nyc_area_df.`

I then grouped each of the data frames `comcast_nyc_area_df,` `verizon_nyc_area_df, time_warner_nyc_area_df,` and `cablevision_nyc_area_df` by `date` and calculated the median download throughput (`daily_median_dtp`) and kept track of the number of measurements (`n`) for each day. Then I graphed

y = daily_median_dtp, x = date