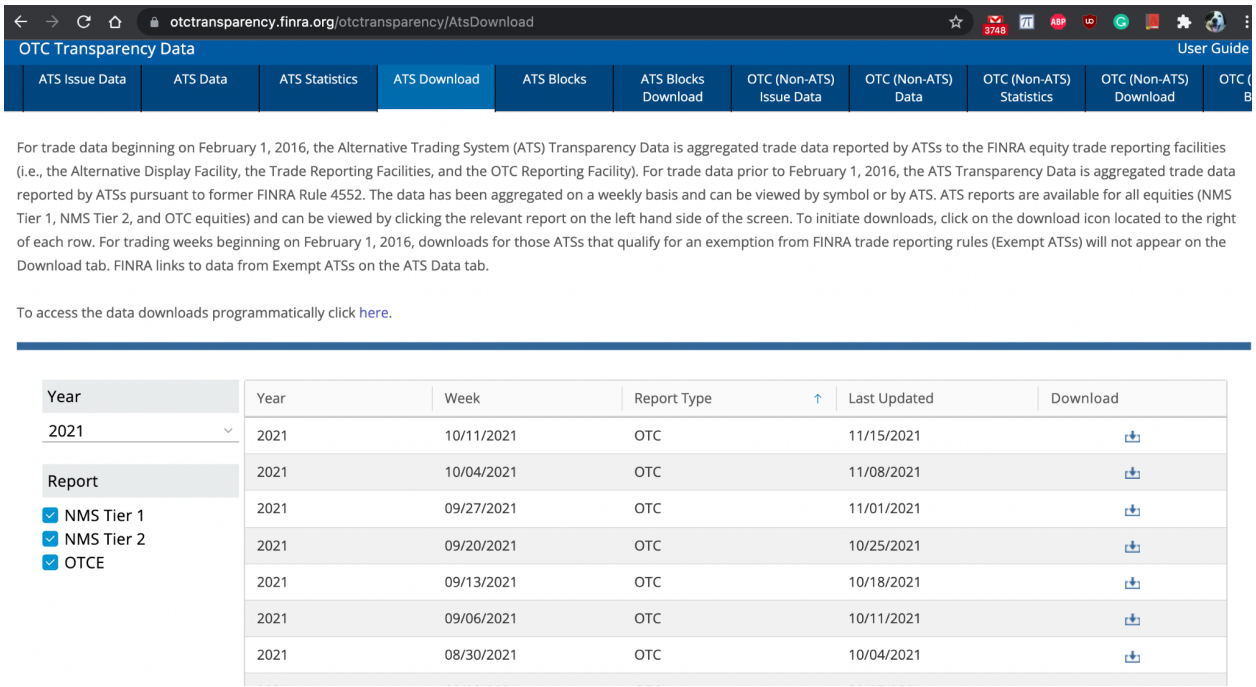
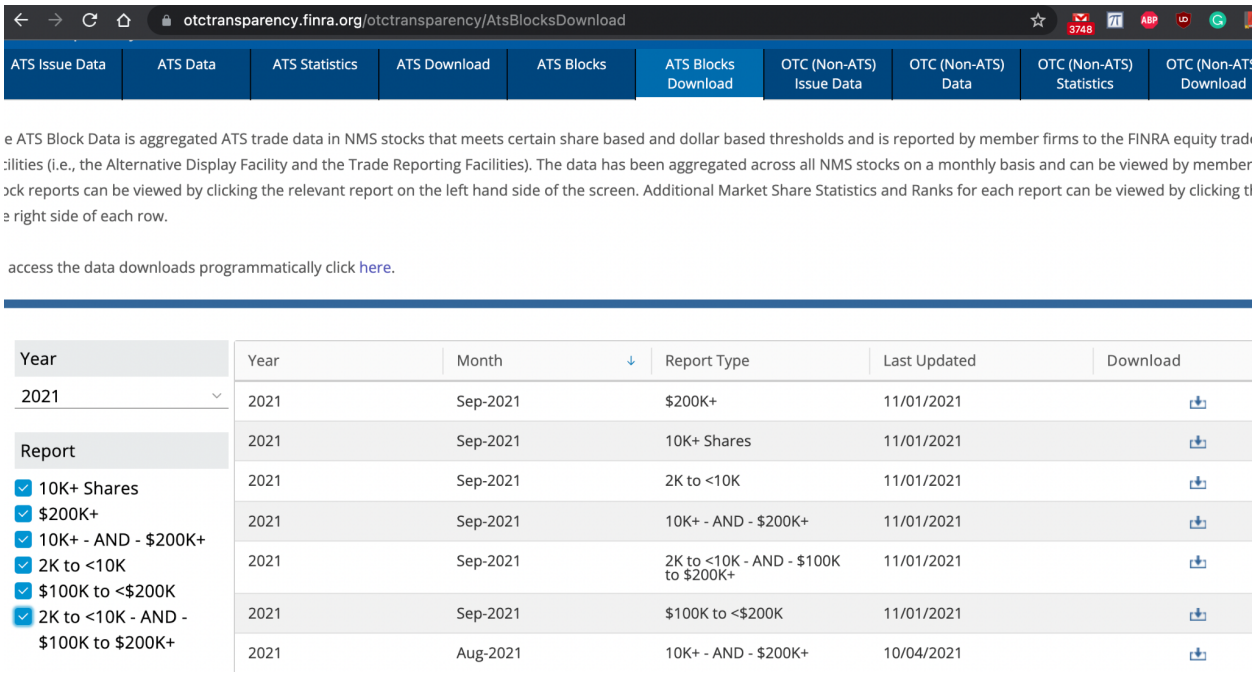
ATS Download

- https://otctransparency.finra.org/otctransparency/AtsDownload
- Enable all three reports
    - NMS Tier 1
    - NMS Tier 2
    - OTCE
- Select the year
- Download all files whose values for Last Updated are after a given reference date (the initial reference date will be prior to the first data set so that all files are downloaded)
    - There are typically 4 or 5 pages per year when all reports are enabled
    - If downloading more than 5 files, employ throttling
    - If downloading more than 10 files, employ throttling and proxy wheel methods
- Parse the resulting pipe delimited html to parquet format
- Label files: ATS_Data_{Report Type}_{Week}.parquet
    - {Report Type} is the value in the Report Type column
        - Replace " " with "_"
    - {Week} is the week formatted in yyyy_MM_dd
- Save files to a given S3 Ingress bucket under a sub-directory with the date the scrape was initiated.



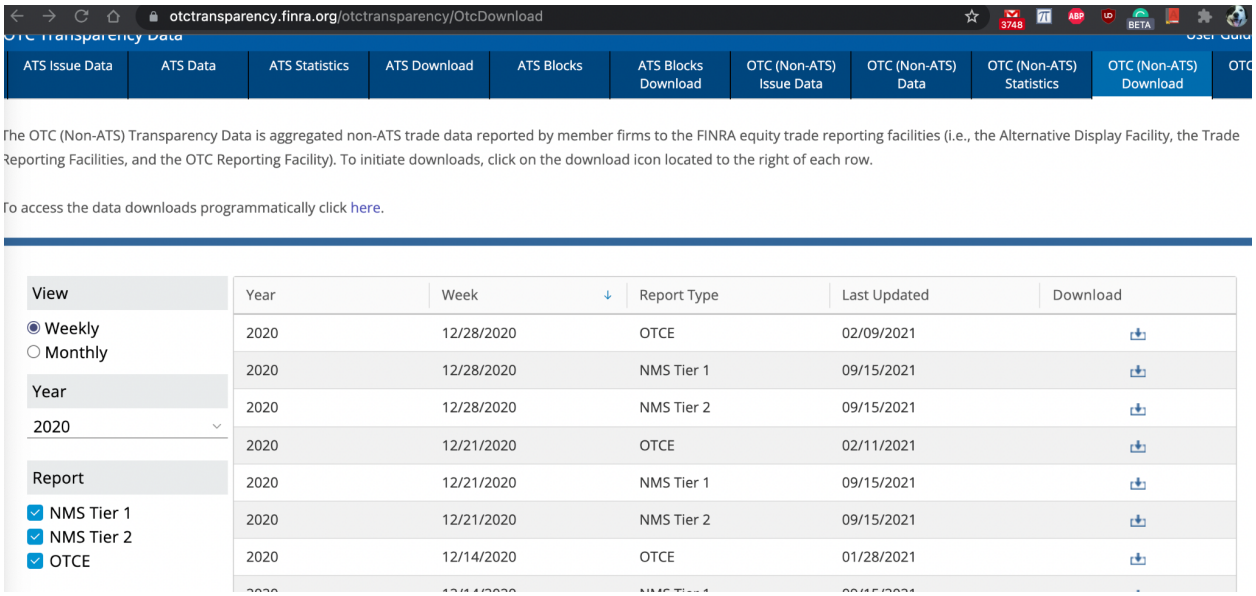
ATS Blocks Download

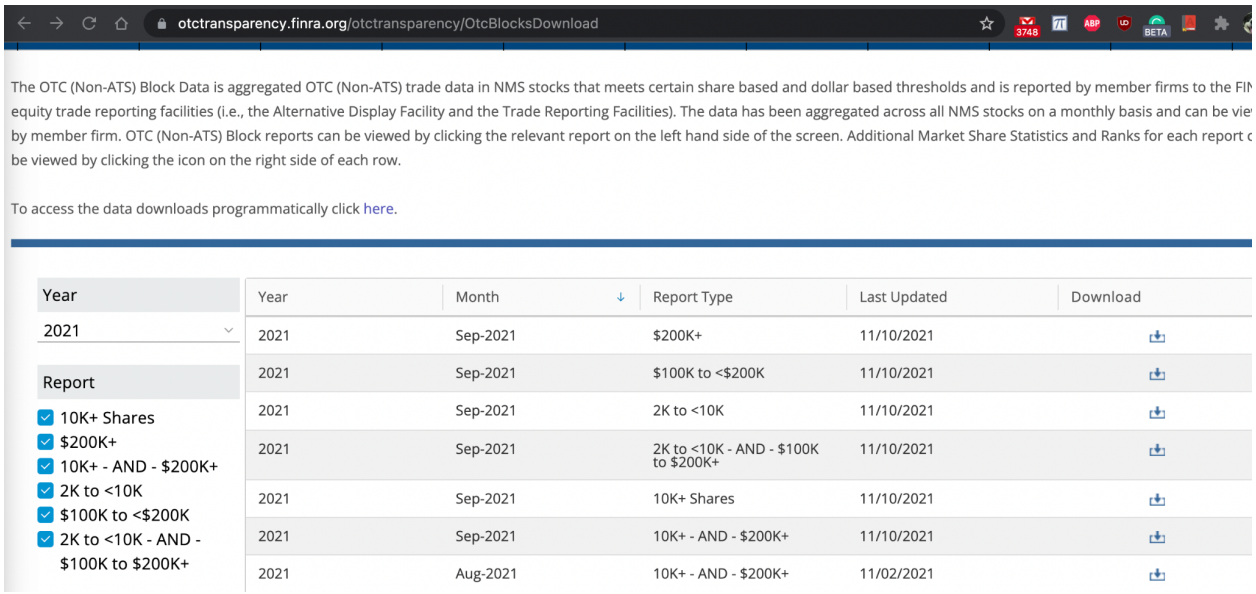
- https://otctransparency.finra.org/otctransparency/AtsBlocksDownload
- Enable all reports
    - 10k+ shares
    - $200k+
    - etc.
- Select the year
- Download all files whose values for Last Updated are after a given reference date (the initial reference date will be prior to the first data set so that all files are downloaded)
    - There are typically 2 pages per year when all reports are enabled
    - If downloading more than 5 files, employ throttling
    - If downloading more than 10 files, employ throttling and proxy wheel methods
- Parse the resulting pipe delimited html to parquet format
- Label files: ATS_Blocks_{Report Type}_{Month}.parquet
    - {Report Type} is the value in the Report Type column.
        - In the following order:
            - Replace “$” with “D”
            - Replace “+” with “p”
            - Replace “<” with “L”
            - Replace “- AND -” with “and”
            - Replace “ “ with “_”
    - {Month} is the Month formatted in yyyy_MM
- Save files to a given S3 Ingress bucket under a sub-directory with the date the scrape was initiated.



OTC (Non-ATS) Download

- https://otctransparency.finra.org/otctransparency/OtcDownload
- Use Weekly View
- Enable all three reports
    - NMS Tier 1
    - NMS Tier 2
    - OTCE
- Select the year
- Download all files whose values for Last Updated are after a given reference date (the initial reference date will be prior to the first data set so that all files are downloaded)
    - There are typically 4 or 5 pages per year when all reports are enabled
    - If downloading more than 5 files, employ throttling
    - If downloading more than 10 files, employ throttling and proxy wheel methods
- Parse the resulting pipe delimited html to parquet format
- Label files: OTC_NonATS_{Report Type}_{Week}.parquet
    - {Report Type} is the value in the Report Type column
        - Replace " " with "_"
    - {Week} is the week formatted in yyyy_MM_dd
- Save files to a given S3 Ingress bucket under a sub-directory with the date the scrape was initiated.



OTC (Non-ATS) Blocks Download

- https://otctransparency.finra.org/otctransparency/OtcBlocksDownload
- Enable all reports
    - 10k+ shares
    - $200k+
    - etc.
- Select the year
- Download all files whose values for Last Updated are after a given reference date (the initial reference date will be prior to the first data set so that all files are downloaded)
    - There are typically 2 pages per year when all reports are enabled
    - If downloading more than 5 files, employ throttling
    - If downloading more than 10 files, employ throttling and proxy wheel methods
- Parse the resulting pipe delimited html to parquet format
- Label files: OTC_NonATS_Blocks_{Report Type}_{Month}.parquet
    - {Report Type} is the value in the Report Type column.
        - In the following order:
            - Replace “$” with “D”
            - Replace “+” with “p”
            - Replace “<” with “L”
            - Replace “- AND -” with “and”
            - Replace “ “ with “_”
    - {Month} is the Month formatted in yyyy_MM
- Save files to a given S3 Ingress bucket under a sub-directory with the date the scrape was initiated.