MonaRk Instructions

Rachel Schweiker

6/10/2020

Installing package

You'll only need to execute this block the first time

```
## ACTION
# 1. Copy MonaRk folder to a location on your computer

# 2. Install Packages
packages <- c("devtools", "shiny", "shinydashboard", "data.table", "DT",
              "ggplot2", "plotly", "colourpicker", "leaflet", "lubridate",
              "sp", "rgeos", "rgdal", "sf", "stringr", "dplyr", "R.utils",
              "svglite", "lwgeom", "shinythemes", "shinycssloaders", "Imap")
if (length(setdiff(packages, rownames(installed.packages()))) > 0) {
  install.packages(setdiff(packages, rownames(installed.packages())))
}
# Sometimes you will get stuck in a loop where R will ask to restart in a popup.
# If you get stuck, say No to restart
```

Change the "/path/to/MonaRk/" and run this block everytime

```
# 3.  change path to MonaRk (including "MonaRk as last directory")
# IN RMARKDOWN
knitr::opts_knit$set(root.dir = "/path/to/MonaRk/")
# OR 3. For copy paste into R session
# setwd("/path/to/MonaRk/")

# 4. Load MonaRk
devtools::load_all()
```

Run App

An instruction video is in the Help Tab of the app.

```
MonaRk::launchApp()
```

Press the stop sign to stop the app and return to R.

Update Data

Use the addMondays() function to update data in the app. This should be run every 2 months to ensure there isn't a lapse in data collection. Data in a usable form resides on the Airbus server under /srv/data/. In order to get it in this form, the directions under the Data Collection and Data Preprocessing sections were followed.

We have a few types of data:

1. Storm Events (polygons representing storms for the USA in 2019).
2. Delay Data (weather and ATC delays for flights in the USA in 2019).
3. Sigmet Data (text that represents airspace dangers released from stations around the world).
4. States Data (the state of the airspace every 10 seconds, Mondays UTC, from OpenSky).

Data Collection

1. Storm Events

The storm events database is released monthly and is 3 months behind the current date.

The most recent storm data was downloaded on 4/18/20 and contains storms through 1/31/20. There is a three-month delay in this data. It will need to be downloaded again to include more recent storms, then reprocessed, and combined with flight data.

```
## ACTION
# 1. Go to https://www1.ncdc.noaa.gov/pub/data/swdi/stormevents/csvfiles/ and download the
# file named "StormEvents_details-ftp_v1.0_d2020_c2020MMDD.csv.gz". The last date will be
# different, but the important thing is that the file name contains "details" and "_d2020_" # OR the cu
# 2. Unzip the gz file
# 3. Replace the 2020 (or whatever year it is currently) storm file in
# MonaRk/inst/extdata/raw/rawStorms with the unziped CSV file you just downloaded
```

2. Delay Data

The delay data is released monthly, with a two or three-month lag (at the end of March 2020, the most recent data available was Jan 2020).

```
# ACTION
# 1. Look in MonaRk/inst/extdata/raw/rawDelay and find the dates at the end of the files. Ex.
# "On_Time_Marketing_Carrier_On_Time_Performance_(Beginning_January_2018)_2020_1.csv", find
# the end "2020_1.csv" which represents January 2020.
# 2. Note which months you'll need to download. Start with the most recent delay file. For
# example, if "2020_1.csv" is the most recent, download January 2020 AGAIN to ensure you have
# the full months worth of data. Download all the months after this date that is possible.
# 3. Go to https://www.transtats.bts.gov/Tables.asp?DB_ID=120 click Download under Marketing.
```

Use "marketing" since it has additional reports over the reporting option.

```
#4. Download one month of data at a time, and (important!) select the "Prezipped File"
# option to get all the fields.
```

Selecting columns sometimes results in the wrong columns being downloaded.

```
#5. Download a file for each month that needs to be updated, unzip them and open them,
# and move the CSV files only into inst/extdata/raw/rawDelay.
```

3. Sigmet Scraping

A script is set to run every hour on an AWS instance which scrapes international sigmets from https://aviationweather.gov/sigmet/intl?&loc=all and US sigmets from https://aviationweather.gov/sigmet/data?hazard=all&loc=all&date=. The result is a folder for each date, each of which contains 10-20 text files representing regions. The names of the files are unimportant.
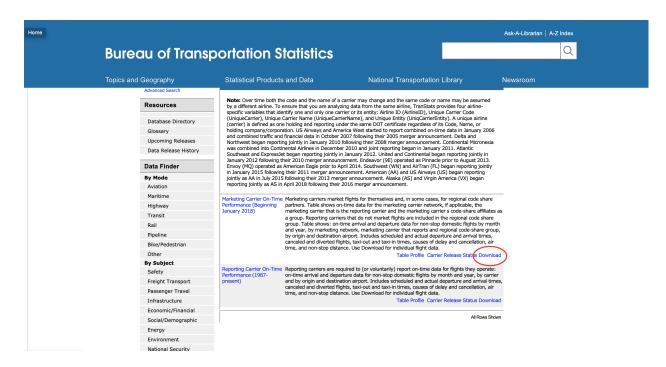
Bureau of Transportation Statistics

Topics and Geography   Statistical Products and Data   National Transportation Library   Newsroom

Advanced Search

Resources

Database Directory
Glossary
Upcoming Releases
Data Release History

Data Finder

By Mode
Aviation
Maritime
Highway
Transit
Rail
Pipeline
Bike/Pedestrian
Other

By Subject
Safety
Freight Transport
Passenger Travel
Infrastructure
Economic/Financial
Social/Demographic
Energy
Environment
National Security

Note: Over time both the code and the name of a carrier may change and the same code or name may be assumed by a different airline. To ensure that you are analyzing data from the same airline, TranStats provides four airline-specific variables that identify one and only one carrier or its entity: Airline ID (AirlineID), Unique Carrier Code (UniqueCarrier), Unique Carrier Name (UniqueCarrierName), and Unique Entity (UniqCarrierEntity). A unique airline (carrier) is defined as one holding and reporting under the same DOT certificate regardless of its Code, Name, or holding company/corporation. US Airways and America West started to report combined on-time data in January 2006 and combined traffic and financial data in October 2007 following their 2005 merger announcement. Delta and Northwest began reporting jointly in January 2010 following their 2008 merger announcement. Continental Micronesia was combined into Continental Airlines in December 2010 and joint reporting began in January 2011. Atlantic Southeast and ExpressJet began reporting jointly in January 2012. United and Continental began reporting jointly in January 2012 following their 2010 merger announcement. Endeavor (9E) operated as Pinnacle prior to August 2013. Envoy (MQ) operated as American Eagle prior to April 2014. Southwest (WN) and AirTran (FL) began reporting jointly in January 2015 following their 2011 merger announcement. American (AA) and US Airways (US) began reporting jointly as AA in July 2015 following their 2013 merger announcement. Alaska (AS) and Virgin America (VX) began reporting jointly as AS in April 2018 following their 2016 merger announcement.

Marketing Carrier On-Time Performance (Beginning January 2018)   Marketing carriers market flights for themselves and, in some cases, for regional code share partners. Table shows on-time data for the marketing carrier network, if applicable, the marketing carrier that is the reporting carrier and the marketing carrier s code-share affiliates as a group. Reporting carriers that do not market flights are included in the regional code share group. Table shows: on-time arrival and departure data for non-stop domestic flights by month and year, by marketing network, marketing carrier that reports and regional code-share group, by origin and destination airport. Includes scheduled and actual departure and arrival times, canceled and diverted flights, taxi-out and taxi-in times, causes of delay and cancellation, air time, and non-stop distance. Use Download for individual flight data.

Table Profile Carrier Release Status Download

Reporting Carrier On-Time Performance (1987-present)   Reporting carriers are required to (or voluntarily) report on-time data for flights they operate: on-time arrival and departure data for non-stop domestic flights by month and year, by carrier and by origin and destination airport. Includes scheduled and actual departure and arrival times, canceled and diverted flights, taxi-out and taxi-in times, causes of delay and cancellation, air time, and non-stop distance. Use Download for individual flight data.

Table Profile Carrier Release Status Download

All Rows Shown

Figure 1: Download Delay Data

On-Time : Marketing Carrier On-Time Performance (Beginning January 2018)

Databases   Data Tables   Table Contents

Download Instructions
Latest Available Data: December 2019

Filter Geography	Filter Year	Filter Period
All	2019	November

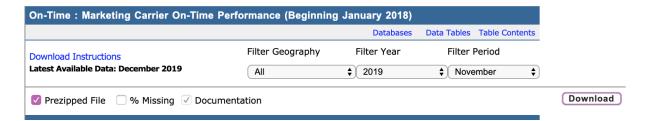☑ Prezipped File   ☐ % Missing   ☑ Documentation

Download

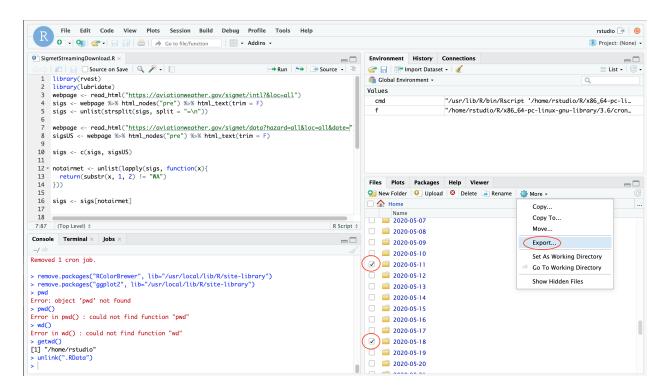Figure 2: Download Delay Data - 2nd page

Figure 3: Download Sigmet Data

```
## ACTION
# 1. Go to http://18.216.88.92/auth-sign-in
# 2. Username: rstudio & Password: i-03765662c5ae43e02
# 3. Look locally in MonaRk/inst/extdata/raw/rawSigmetScraping to see what the latest
# sigmet date is that has been downloaded
# 4. Export all Mondays since this date from the RStudio server
```

```
## ACTION CONTINUED
# 5. The name of the zipped download doesn't matter, just unzip and select the date
# folders and move them into MonaRk/inst/extdata/raw/rawSigmetScraping
```

Sidenote: Sigmets from before 02-09-2020 are from http://aviationwxchartsarchive.com/product/sigmet. Not as many sigmets are captured by this service so we switched to scraping our own.

4. States Data

Flight data is hosted on https://opensky-network.org/datasets/states/. Inside the date folders are folders for each hour which contains a csv.tar file. You can download and unzip these manually, or use the function readOpenSkyStates() to run through a day automatically.

Raw states data is stored on the Airbus server in /srv/data/flight_states since it is too large to be included in the MonaRk package files.

```
# ACTION
# 1. Proceed to the next step and addMondays() will download states data automatically.
```

Add Mondays function

You will need a large amount of memory to run this function. For each week that has passed since updating the data, this function will download about 3GB worth of data.

```
## ACTION: change path for data download and app folder
appDataDir <- "path/to/MonaRk"
statesDataDir <- "path/to/download"

addMondays(appDataDir, statesDataDir)
```

If you've forgotten to update any of the simget, storm event, or delay data, an error will tell you which data to download into which folder.

After running the addMondays function

Many large states files will likely be downloaded to the Directory you defined as "statesDataDir". Once addMondays() has finished processing, you can navigate to "statesDataDir" and remove any files with names in the format states_YYYY-MM-DD.csv. It is recommended to move these files so the Airbus server: /srv/data/flight_states in case these files need to be reprocessed.

Check out the app with the new data

```
MonaRk::launchApp()
```

Does the data appear to be updated? Does the date selection on the last tab include recent dates? If something went wrong, the data before updating will be stored in MonaRk/inst/extdata/cleaned with the suffix _old.RDS. You can restore the old data by deleting any .RDS files in MonaRk/inst/extdata/cleaned without the suffix _old.RDS, and removing "_old.RDS" from the names of the files. The app then will use the old data files.

Updating the online app

If the code for the app is changed, or if you'd like to distribute the app with updated data, you can publish a new version of the app on shinyapps.io. You can find a guide here: https://towardsdatascience.com/how-to-publish-a-shiny-app-example-with-shinyapps-io-ec6c6604d8. ONLY push 1. "app.R", 2. the "www" folder, 3. the "appData" folder, and 4. the "R" folder. The setup should take ~15 minutes and the publishing itself takes a few hours because of large data files.

Conducting your own analysis

The parsed data could be helpful for many other projects. 1. Parsed US storm events data exists in: MonaRk/inst/extdata/cleaned/stormsGeo.RDS 2. Parsed US delay data exists in: MonaRk/inst/extdata/cleaned/delayGeo.RDS 3. Parsed sigmet data exists in: MonaRk/appData/allsigs.RDS 4. Parsed flight data exists as two files. Read it into R with the following code:

```
flights <- get(load("inst/extdata/cleaned/df.RDS"))
geo <- get(load("inst/extdata/cleaned/geo.RDS"))

flightgeo <- createFlightgeo(flights, geo)
```