This is the README file for the second stage code of the article "Monitoring War Destruction from Space Using Machine Learning"

In order to replicate the results first run the jupyter notebook “single\_second\_stage\_replication” in the replication folder “replication\_second\_stage\_tandt” which also contains some documentation of the code. The code produces Table 1 and Table 2 together with all other performance statistics like ROC curves and precision/recall curves. It also produces a file called “performance\_stats” which contains more statistics from the paper.

The Stata do-file “second\_stage\_output.do” produces the event study regression output Table S2 (SI) and the coefficient plot Figure 5. It combines the second-stage prediction output with the LiveUAmap event data.

The program was written in Stata 14.2 and requires installing the following additional packages:

i) coefplot (<https://ideas.repec.org/c/boc/bocode/s457686.html>) developed by Jann (2014),

ii) outreg2 (<http://repec.org/bocode/o/outreg2.html>) developed by Wada (2005), and

iii) colorpalettes (<https://github.com/benjann/palettes/>) developed by Jann (2018).

The resulting dataset can also be used for reproducing Figure 3c and 4. Note that the figures were produced manually in QGIS 3.1 and are not automatically generated by the program. Details for reproduction are provided below.

**Figure 3c**: depicts the average spatial distribution of the variable “random\_forest\_pred” (i.e. the second stage dense patch-wise destruction prediction scores) for Aleppo city. Color bins reflect deciles of the distribution of this variable. Green color indicates low prediction scores, red color indicates high prediction scores.

**Figure 4**: depicts the spatial distribution of the continuous variable “random\_forest\_pred” (i.e. the second stage dense patch-wise destruction prediction scores) and the binary variable “pred\_binary” (i.e. the binary prediction based on “random\_forest\_pred” for an arbitrary cutoff optimized to reach 50 percent recall in the test sample) across a selected non-analysis area in the Ramouse neighborhood of Aleppo (approximate image centroid location: 36.1525 decimal degrees North and 37.1332 East). The two time periods depicted are 12th of June 2016 (pre-image corresponding to wave 14) and 18th of September 2016 (post-image corresponding to wave 17). For the continuous destruction prediction in panels b) and e), color bins reflect deciles of the aggregate distribution in Aleppo for the predictions of the image on 18th of September 2016. For the binary version in panels c) and f), green reflects “undestroyed” and red “destroyed”.

**References**

Jann, Ben (2014). Plotting regression coefficients and other estimates. The Stata Journal. Volume 14, Number 4: pp. 708-737.

Wada, Roy (2005). OUTREG2: Stata module to arrange regression outputs into an illustrative table. Statistical Software Components S456416, Boston College Department of Economics, revised 17 Aug 2014.

Jann, Ben (2018). Color palettes for Stata graphics. The Stata Journal. Volume 18, Number 4: pp. 765-785.