Data and Replication Information for "Corruption Dynamics: The Golden Goose Effect."
Paul Niehaus (UCSD) and Sandip Sukhtankar (Dartmouth College)

The following files were used to create the tables and analysis presented in the paper. All analysis was conducted using R. Comments in the code describe details of the analysis.

1. *finalanalysisdata.csv* and *regressions [FINAL].r*
These are the data file and code respective that create most of the tables and figures (other than exceptions noted in 2 and 3 below). Each observation in the data file corresponds to a panchayat-day. Descriptions for each of the variables in the data are included in the code file. The sources of raw data as well as intermediate datasets are described in the paper and appendix. In order to use these files to recreate paper tables, you will also need the following supplemental code files:

    i.   [5-5-2008][v2] distributor_functions.R. This file contains contains functions converting a given date into day of year. Also contains functions that distribute amounts earned and days worked evenly between a given period when the exact days worked are not known.

    ii.   lib_mwc.r. This file does multi-way clustering a la Cameron-Gelbach-Miller (2011)

    iii.   lib_multiregtable.r. Helps with outputting tables into LaTeX.

    iv.   lib_hacktex.r. Also helps with outputting tables into LaTeX.

2. *didanalysisdata.csv* and *regressions_did [FINAL].r*
These data and code files produce the difference-in-difference tables (Table 5 and C4). Each observation in the data file corresponds to a panchayat-day. Descriptions for each of the variables in the data are included in the code file. The sources of raw data as well as intermediate datasets are described in the paper and appendix. The same supplemental code files described above are also used in this instance.

3. *village_data_final.csv* and *audit_MLE_estimation [FINAL].R*
These data and code files produce the table on audit probabilities post shock (Table 8). Each observation in the data file corresponds to a panchayat. The main outcome variables are the date and month of visit to the panchayat by the Block Development Officer (BDO) and District Collector. The estimation strategy is described in the paper.