World Bank Geocoded Dataset
1.0 Release: v2015/01/20

Please direct any questions or comments to data@aiddata.org

Geocoded Data Coverage

Years: 2000 - 2011

Donors: 1

Total Projects: 3578

Geocoded Projects: 3534

Locations: 41307

Total Disbursements: $0.00

Total Commitments: $369,682,225,142.57

Geocoded Disbursements: $0.00

Geocoded Commitments: $366,835,699,448.50

Glossary

ADM levels: sub-national administrative areas as of 2011

AidData: a partnership between the College of William and Mary, BYU, and DG

AMP: Aid Management Platform, made by DG

DG: Development Gateway, a nonprofit organization

Files Included in this Release

File or Folder	File or Sub-Folder	Description
	README_WBK.pdf	This file
data	projects.tsv	World Bank projects table
data	locations.tsv	locations for each project
data	transactions.tsv	transactions for each project
data	ancillary_wb_projects.tsv	ancillary data from WB projects database (not quality assured)
data	ancillary_wb_evaluations.tsv	ancillary data from evaluations database (not quality assured)
metadata	country_codes.tsv	country name to iso2 crosswalk.
metadata	data_dictionary.tsv	field name and description in the data tables.
metadata	deflator_values.tsv	deflator table used for deflation of financial amounts.
metadata	donor_trans_crosswalk.tsv	donor crosswalk file
metadata	geonames.tsv	geonames valid location types.

metadata	sector_trans_crosswalk.tsv	sector translation/crosswalk file.
doc	ucdp_aiddata_codebook_published.pdf	UCDP Codebook (geocoding methodology)

Overview

This geocoded dataset release represents all World Bank projects in the IBRD and IDA lending lines approved from 2000-2011. Each record in the projects table is associated with at least one record in the locations table. The files project.tsv, locations.tsv, and transactions.tsv are structured so as to be combinable with other v1.0 geocoded dataset releases.The file ancillary_wb_projects.tsv contains all fields that are found in exports from the World Bank project database.(http://www.worldbank.org/projects/advancedsearch?lang=en). The file ancillary_wb_evaluations.tsv contains all IEG evaluations that have been conducted on World Bank projects approved from 2000-2011 (http://data.worldbank.org/data-catalog/IEG). All tables in the data folder can be joined by unique World Bank project id.

Using Data in a GIS

AidData has developed an Introduction to ArcGIS Training Module (https://drive.google.com/folderview?id=0Bysb0tDCQWQxd05LQUstWTlRTjQ&usp;=sharing) . This training is oriented towards new users of spatial information with the purpose of improving understanding and ability to use spatial information. The instructional tutorial will help users to join spatial information with other data, calculate new variables in ArcGIS, and create cartographic outputs.

Using Data in Spreadsheets

Microsoft Excel has known limitations when working with TSV and CSV files that have character encodings like UTF8 (the character encoding of the AidData CSV Data).To avoid potential problems with character encodings, please use the "From Text"command on the Data menu when opening the CSV files. Double-clicking on the files to open in Excel, and/or saving them as CSV from Excel may result in character encoding problems.

For instructions, see this link: http://office.microsoft.com/enus/excelhelp/importorexporttexttxtorcsvfilesHP010099725.aspx#BMimport _data_from_a_text_file_by_openi

Additionally, AidData provides all final tabular data as TAB separated files. If asked by your software which delimiter to use, you should select TAB. **PLEASE BE CAREFUL TO ONLY USE ONE DELIMITER (not for example, comma and TAB)**

The four (4) files in the research release are related as follows:
1) The projects.tsv table is the main table. It contains one row per project. The unique id is the project_id field.
2) The locations.tsv table is the locations table. It contains one row per location. It is related to the projects table by the project_id fields in each table. Each project may have many locations. The unique id is the project_location_id field.
3) The transactions.tsv table is the transactions (financials) table. It contains one row per transaction. It is related to the projects table by the project_id fields in each table. Each project may have many transactions. There is no unique field for the transactions table.
4) The ancillary.tsv table contains all ancillary data for the projects table. It contains any ancillary data for the project table. It is related to the projects table by the project_id fields in each table. Each project may have many ancillary records. There is no unique field for the ancillary table.

Caveats

*Because of licensing restrictions, we cannot include the GIS files for these boundaries along with the release. Please visit http://www.fao.org/geonetwork/srv/en/metadata.show?id=12691 for these files.

*This dataset was not activity coded. AidData sector codes are crosswalked from the sector designations in the World Bank database. See crosswalk table in the metadata folder.

*Financials are deflated to USD2011, with the exception of transactions >= 2013, which are reported in current USD

*MJSECTOR 1 was cross-walked to AidData Purpose Codes. No other sector/purpose/activity code cross-walking was performed

*13 World Bank projects were identified as being cancelled in their documentation

*Projects locations of precision code 7 (Unknown Location) were not geocoded, so there will be now precision code 7 locations in the dataset. The projects themselves are still included.

*No disbursement data was used for geocoding (please see the $0 disbursement figure in the summary statistics)

Precision Code

Precision Code Value	Precision Code Description
1	coordinates correspond to an exact location
2	coordinates correspond to a district or a location that is known to be within 25km of the coordinates
3	coordinates correspond to a zone
4	coordinates correspond to a region
5	estimated coordinates of a large feature, such as rivers or national parks
6	coordinates correspond to the entire country, project operates in sub-national locales but they are not known
8	coordinates correspond to the entire country, it is likely that the funding goes to a government ministry or financial institution

Data Coverage Charts

See charts below. The first chart (Chart A) breaks down the entire dataset by precision code. The second chart (Chart B) breaks down the geocoded data by sector and precision code. The third chart (Chart C) charts the number of projects by start years.
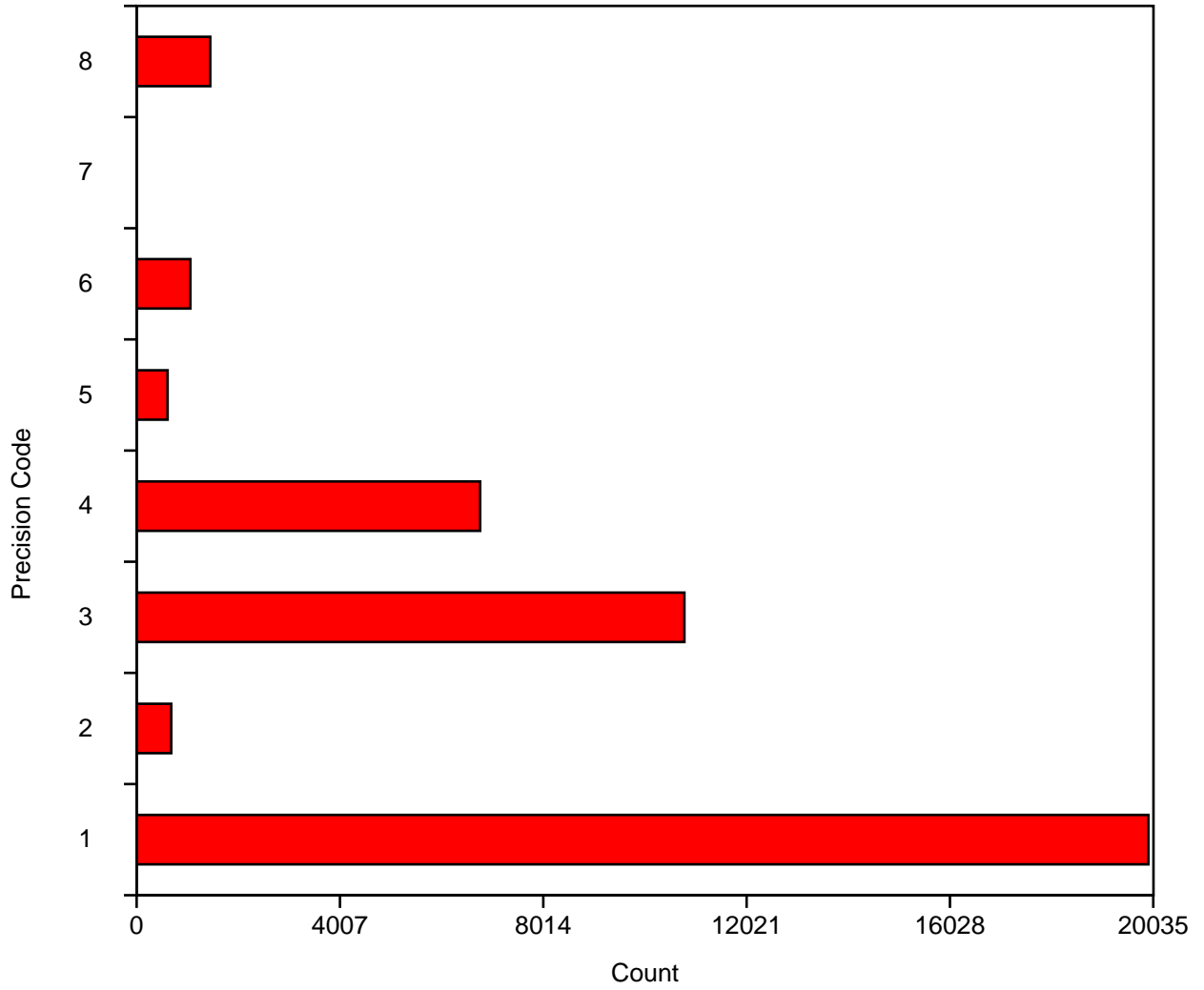
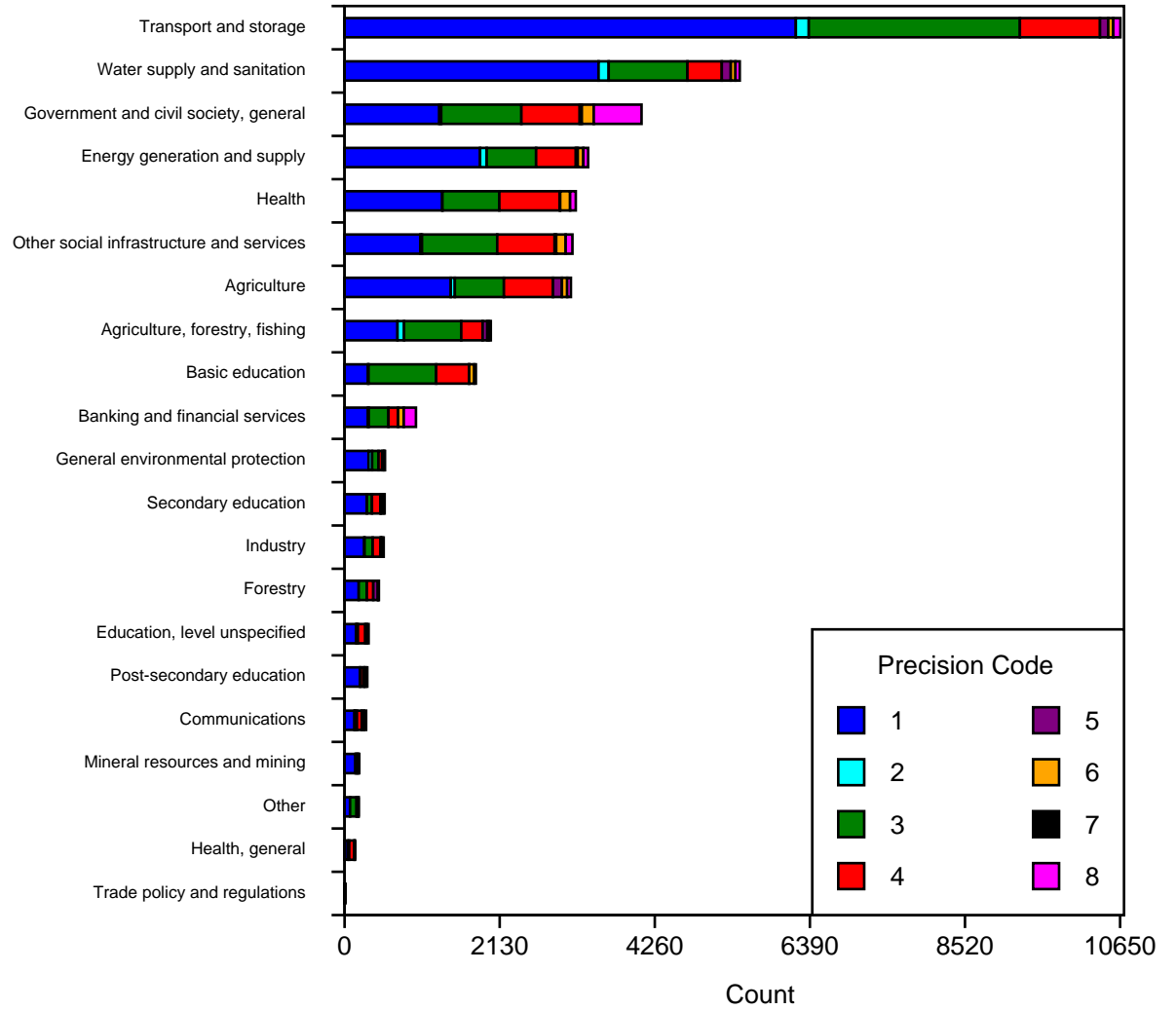**Chart A, Precision Codes**
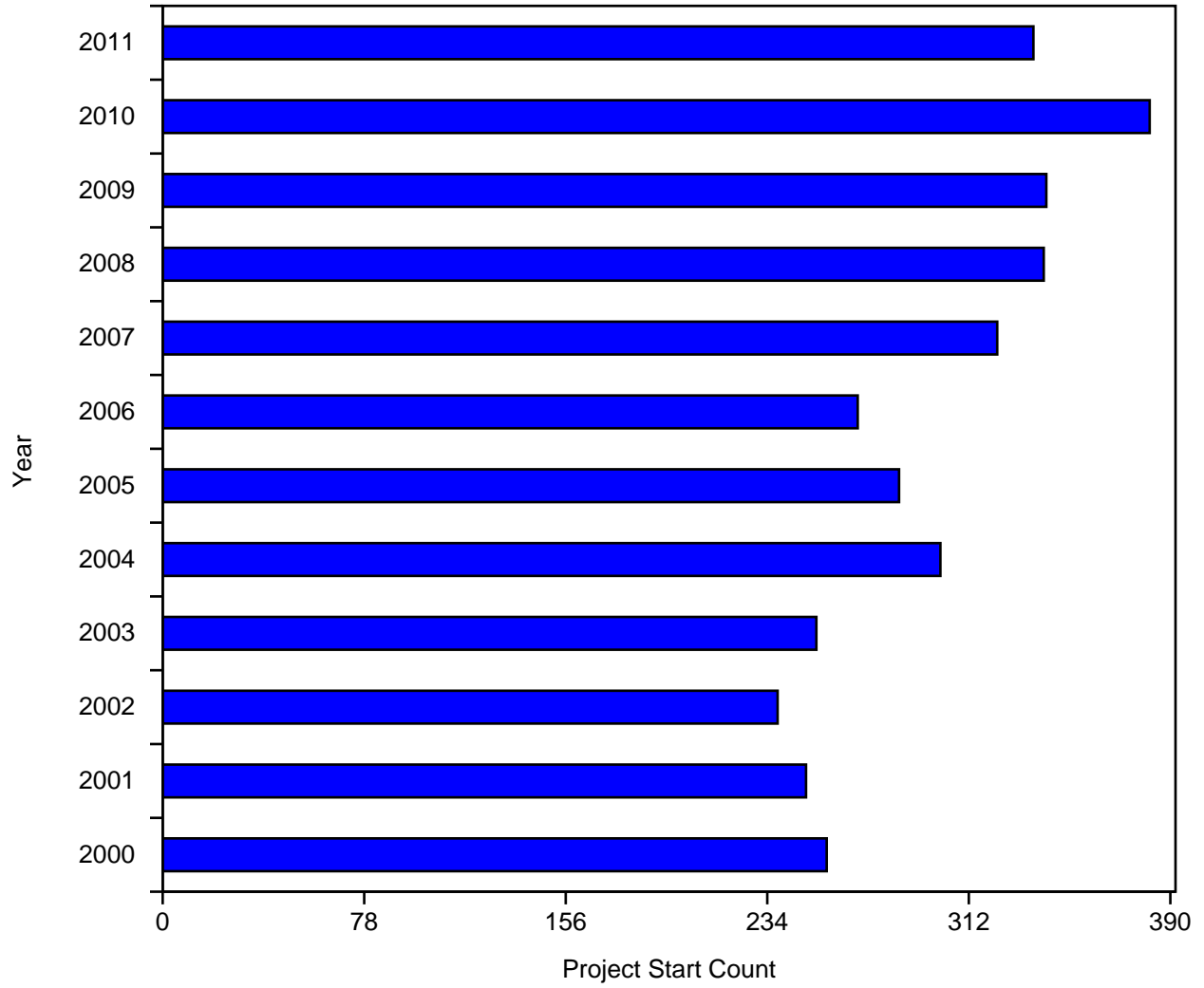
Chart B, Precision Codes by Sector

Precision Code

1	5
2	6
3	7
4	8

Count

Chart C, Project Start Counts By Year

Websites of Sources

Source	URL
AidData	http://aiddata.org/
GeoNames	http://www.geonames.org/
GeoNames API Example	http://ws.geonames.org/get?geonameId=232422&style=full
GeoNames Gazetteer	http://download.geonames.org/export/dump/
IATI Organization (Donor) Identifiers	http://iatistandard.org/codelist/organisation/
Boundary Shapefiles	Not Provided (see Caveats)
World Bank Project Database	http://www.worldbank.org/projects/advancedsearch?lang=en
IEG Rating	http://data.worldbank.org/data-catalog/IEG