



文献 CSTR:

32001.14.11-6035.csdata.2021.0034.zh

文献 DOI:

10.11922/csdata.2021.0034.zh

数据 DOI:

10.11922/sciencedb.j00001.00273

文献分类: 社会科学

收稿日期: 2021-06-02

开放同评: 2021-08-13

录用日期: 2022-06-08

发表日期: 2022-12-23

* 论文通信作者

高亮: lianggao@bjtu.edu.cn

全国六次人口普查区县级人口分布数据集

吴昊然¹, 高亮^{1*}, 宋栋栋¹, 杨一涛¹, 徐长兴¹, 杨小宝¹

1. 北京交通大学, 综合交通运输大数据应用技术交通运输行业重点实验室, 北京 100044

摘要: 人口空间化数据对于支持交通、地理、社会经济、可持续发展等方面的研究具有重要意义。本研究搜集了 1949 年以来中国六次人口普查的区县级人口统计数据。以 2015 年 (第六次人口普查年) 区县级行政区划为区县行政区的地理区域基准, 对人口统计数据进行空间化校准, 经空间化处理, 生成区县级 GIS 人口数据集。

关键词: 人口普查; 空间化; 区县; 中国

数据集基本信息简介

数据集名称	全国六次人口普查区县级人口分布数据集
数据作者	吴昊然、高亮、宋栋栋、杨一涛、徐长兴、杨小宝
数据通信作者	高亮 (lianggao@bjtu.edu.cn)
数据时间范围	1953年, 1964年, 1982年, 1990年, 2000年, 2010年
地理区域	中国大陆
数据量	67.6 MB
数据格式	*.shp, *.cpg, *.dbf, *.prj, *.sbn, *.sbx, *.shx
数据服务系统网址	http://www.doi.org/10.11922/sciencedb.j00001.00273
基金项目	国家自然科学基金 (71571017, 91646124, 71621001, 91746201)
数据库 (集) 组成	本数据集包含“县区_POP_20210602.rar”包含经过编码、属性连接等形成中国31省 (自治区、直辖市) (香港、澳门、台湾省除外) 区县级GIS人口数据集。

引言

人口数据为城市规划^[1]、交通^[2-4]、气候与环境^[5-6]、地理^[2,7]、社会科学^[3,8-9]、经济与管理^[3,8-9]等相关科学领域提供重要的基础数据支撑, 对这些学科的发展具有重要意义。近年来, 人口数据生产也得到了国内外研究者的重视^[10-15]。

从新中国成立到 2010, 我国开展了六次全国人口普查, 涉及全国 31 个省 (自治区、直辖市)。六次人口普查, 投入人力、物力、财力充足, 人口统计数据精度较高。以此为基础, 生成空间化人口数据, 数据质量较高。因此, 本研究采集六次人口普查年对应的人口统计数据, 经过数据校准和空间化处理, 得到了“中国区县级人口普查 GIS 数据集 (1953–2010)”。

已有中国人口数据生产，主要采用网格/栅格划分全国的地理区域。为了获取每个网格的人口数据，需要对高级别行政区域人口数据，进行必要的变换，最终输出数据产品^[10,12-13,15]。本研究以实际区县级行政区域划分全国的地理区域，每个区县级行政区在人口普查中，都有对应的人口数据，不需要进行数据变换，避免了数据变换过程中产生的系统性数据误差。另外，以区县级行政区为生产人口数据的基本地理单元，可以较好地反映人们的出行或迁移等相关特征，提高相关研究结果的准确性和应用性。

1 数据采集和处理方法

1.1 数据来源

本数据集的核心数据源为中国六次人口普查的人口数据（香港、澳门、台湾省除外）。在数据集制作过程中，具体用到的数据集包括：国家统计局发布的区县级人口统计数据（<http://www.stats.gov.cn>）；2015 年全国区县级行政区划 GIS 矢量数据（<http://www.resdc.cn>），其地理参考为 WGS84 坐标系。

1.2 数据处理

相对于 2015 年的行政区划，我国前五次人口普查年对应的行政区划发生了变化。包括区县更名、行政区等级或属性发生改变，以及行政区的合并或拆分。对于上述变化，以 2015 年全国区县级行政区划为基准，对前五次人口普查年对应的行政区划矢量数据进行数据校准，得到行政区划校准数据。

接着，将对应的人口统计数据与区县级行政区进行匹配。例如，1953 年某个行政区，2015 年时已经被拆分，假设 1953 年该行政区内人口分布是均匀的，那么，按 2015 年拆分后所得各行政区面积比例，将 1953 年该行政区人口数据拆分到 2015 年新行政区中。

最后，以 2015 年行政区划为基准，在 2015 年区县行政区 GIS 数据基础上，将人口数据进行空间化处理，得到“全国六次人口普查区县级人口分布数据集”。

综上，本数据集生产技术路线如图 1 所示。

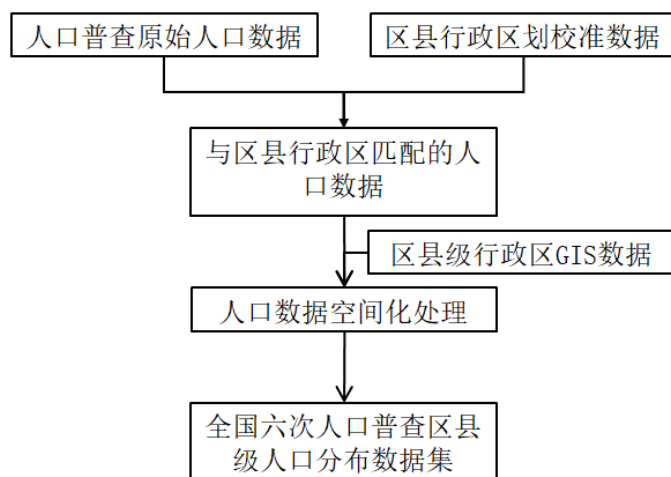


图 1 数据集生产技术路线图

Figure 1 Technology roadmap of the dataset production

关于区县级人口数据空间化处理方法，具体如下：

(1) 区县级行政区划数据校准。本研究所生产的数据集以 2015 年区县级行政区划为基准，对 1953–2010 年六次全国人口普查时，区县级行政区划进行校准。其中，根据区县级行政区的调整，合并或拆分了一些区县级行政区。

(2) 人口普查数据的行政区划校准。假设各区县级行政区内人口是均匀分布的，生成相对于 2015 年行政区划发生了调整的区县级行政的人口数据。

(3) 人口普查数据的空间化处理。本研究获取中国科学院资源环境科学数据中心提供的 2015 年全国区县级行政区划矢量数据集 (<http://www.resdc.cn>)。将六次全国人口普查数据与 2015 年全国区县级行政区矢量数据，建立关联，给每个区县级行政区添加 6 个属性数据，分别对应于 6 次人口普查年的人口。最终，得到中国 31 省（自治区、直辖市）2866 个区县级行政区的人口空间数据集。

2 数据样本描述

经过处理，中国 31 省（自治区、直辖市）区县人口分布如图 2 所示。

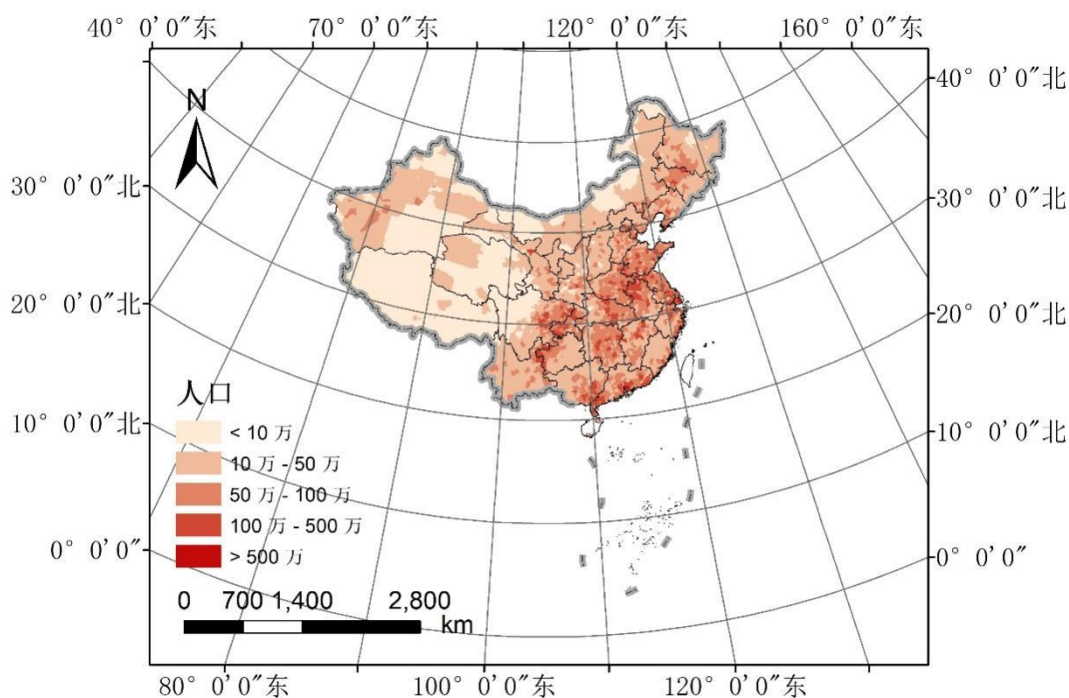


图 2 2010 年第六次人口普查人口分布图（无香港、澳门、台湾数据，审图号：GS(2022)981 号）

Figure 2 Population distribution map of the Sixth Census in 2010 (No data for Hong Kong, Macao and Taiwan)

3 数据质量控制和评估

主要通过以下手段对数据集进行质量控制：

(1) 2015 年区县级行政区矢量数据，来源于中国科学院资源环境科学数据中心，数据来源真实可靠，数据格式、内容规范。

(2) 六次人口普查数据来源于中国国家统计局, 公布的人口普查年(1953 年、1964 年、1982 年、1990 年、2000 年、2010 年) 的统计年鉴, 数据来源可靠, 数据内容规范。

(3) 人口统计数据与区县级行政区进行地理空间匹配过程中, 对发生了拆分的行政区, 假设其人口空间分布均匀, 进而以区域面积比例分配人口量。

由于六次人口普查数据中, 海南省部分区县人口数据缺失, 图 2 中显示为空白。

4 数据使用方法和建议

中国 31 省(自治区、直辖市)(香港、澳门、台湾省除外) 六次人口普查区县级 GIS 人口数据集, 可用 ArcGIS 等地理信息系统软件打开。本数据集可直接用于我国人口分布的时空演化特性研究。同时, 本数据集可进一步支持交通、地理、社会科学、经济与管理等方面与人口迁移演化相关的课题研究。本课题组正基于本数据集, 研究全国范围内交通系统, 随人口空间分布变化的演化规律, 欢迎感兴趣的课题组, 参与深入讨论。

致 谢

感谢北京交通大学 2019 年春季大系统理论课上参与原始数据采集的全体同学。

数据作者分工职责

吴昊然(1996—), 男, 广西柳州人, 硕士研究生, 研究方向为交通复杂网络。主要承担工作: 数据集制作, 样本数据分析。

高亮(1979—), 男, 黑龙江齐齐哈尔人, 博士, 副教授, 研究方向为复杂网络、复杂系统管理、交通系统复杂性。主要承担工作: 课题设计, 数据集制作, 论文撰写, 样本数据分析。

宋栋栋(1995—), 男, 甘肃天水人, 博士研究生, 研究方向为交通系统复杂性。主要承担工作: 数据集制作, 论文撰写, 样本数据分析。

杨一涛(1995—), 男, 安徽合肥人, 博士研究生, 研究方向为交通系统复杂性。主要承担工作: 数据集制作。

徐长兴(1990—), 男, 山东菏泽人, 硕士研究生, 研究方向为交通系统复杂性。主要承担工作: 数据集制作。

杨小宝(1981—), 男, 江西吉安人, 博士, 教授, 研究方向为交通系统复杂性、交通行为与安全。主要承担工作: 论文撰写。

参考文献

- [1] FANG Y, JAWITZ J W. The evolution of human population distance to water in the USA from 1790 to 2010[J]. Nature Communications, 2019, 10: 430. DOI: 10.1038/s41467-019-08366-z.
- [2] 张君, 薛亮, 延军平. 陕北黄土高原人类活动价值度变化分析[J]. 干旱区资源与环境, 2020, 34(9): 1–8. DOI: 10.13448/j.cnki.jalre.2020.232. [ZHANG J, XUE L, YAN J P. Change of human activity value degree in Loess Plateau of northern Shaanxi[J]. Journal of Arid Land Resources and Environment, 2020,

34(9): 1–8. DOI: 10.13448/j.cnki.jalre.2020.232.]

[3] FLÜCKIGER M, LUDWIG M. Transport infrastructure, growth and persistence: the rise and demise of the Sui Canal[J]. *Canadian Journal of Economics*, 2019, 52(2): 624–666. DOI: 10.1111/caje.12378.

[4] BU C J, CUI X Q, LI R Y, et al. Achieving net-zero emissions in China's passenger transport sector through regionally tailored mitigation strategies[J]. *Applied Energy*, 2021, 284: 116265. DOI: 10.1016/j.apenergy.2020.116265.

[5] MANOLI G, FATICHI S, SCHLÄPFER M, et al. Magnitude of urban heat Islands largely explained by climate and population[J]. *Nature*, 2019, 573(7772): 55–60. DOI: 10.1038/s41586-019-1512-9.

[6] LI D J, STUCKY B J, DECK J, et al. The effect of urbanization on plant phenology depends on regional temperature[J]. *Nature Ecology & Evolution*, 2019, 3(12): 1661–1667. DOI: 10.1038/s41559-019-1004-1.

[7] 柏中强, 王卷乐, 杨雅萍, 等. 基于乡镇尺度的中国 25 省区人口分布特征及影响因素[J]. *地理学报*, 2015, 70(8): 1229–1242. DOI: 10.11821/dlxb201508004. [BAI Z Q, WANG J L, YANG Y P, et al. Characterizing spatial patterns of population distribution at township level across the 25 provinces in China[J]. *Acta Geographica Sinica*, 2015, 70(8): 1229–1242. DOI: 10.11821/dlxb201508004.]

[8] GAO J, ZHANG Y C, ZHOU T. Computational socioeconomics[J]. *Physics Reports*, 2019, 817: 1–104. DOI: 10.1016/j.physrep.2019.05.002.

[9] DONG L, RATTI C, ZHENG S Q. Predicting neighborhoods' socioeconomic attributes using restaurant data[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2019, 116(31): 15447–15452. DOI: 10.1073/pnas.1903064116.

[10] 柏中强, 王卷乐. 中国 27 省乡镇(街道)级人口密度数据集(2000 年)[J]. *中国科学数据*, 2016, 1(1): 37–42. DOI: 10.11922/csdata.170.2015.0002. [BAI Z Q, WANG J L. A dataset of population density at township level for 27 provinces of China(2000)[J]. *China Scientific Data*, 2016, 1(1): 37–42. DOI: 10.11922/csdata.170.2015.0002.]

[11] 王明明, 王卷乐. 山东省街道/乡镇级格网化人口空间分布数据集(2000、2010 年)[J]. *中国科学数据*, 2019, 4(3). (2019-04-18). DOI: 10.11922/csdata.2019.0005.zh. [WANG M M, WANG J L. Gridded data of township-level population distribution in Shandong Province(2000 & 2010)[J]. *China Scientific Data*, 2019, 4(3). (2019-04-18). DOI: 10.11922/csdata.2019.0005.zh.]

[12] 王卷乐, 王晓洁, 王明明, 等. 中国乡镇(街道)人口密度数据集(2010 年)[J]. *中国科学数据*, 2020, 5(3). (2020-09-11). DOI: 10.11922/csdata.2020.0013.zh. [WANG J L, WANG X J, WANG M M, et al. A dataset of township-level population density in China(2010)[J]. *China Scientific Data*, 2020, 5(3). (2020-09-11). DOI: 10.11922/csdata.2020.0013.zh.]

[13] GAUGHAN A E, STEVENS F R, HUANG Z J, et al. Spatiotemporal patterns of population in mainland China, 1990 to 2010[J]. *Scientific Data*, 2016, 3: 160005. DOI: 10.1038/sdata.2016.5.

[14] MEYER M F, LABOU S G, CRAMER A N, et al. The global lake area, climate, and population dataset[J]. *Scientific Data*, 2020, 7: 174. DOI: 10.1038/s41597-020-0517-4.

[15] CHEN Y D, GUO F, WANG J C, et al. Provincial and gridded population projection for China under shared socioeconomic pathways from 2010 to 2100[J]. *Scientific Data*, 2020, 7: 83. DOI: 10.1038/s41597-020-0421-y.

论文引用格式

吴昊然, 高亮, 宋栋栋, 杨一涛, 等. 全国六次人口普查区县级人口分布数据集[J/OL]. 中国科学数据, 2022, 7(4). (2022-06-08). DOI: 10.11922/csdata.2021.0034.zh.

数据引用格式

高亮, 吴昊然, 宋栋栋, 杨一涛, 等. 全国六次人口普查区县级人口分布数据集[DS/OL]. Science Data Bank, 2021. (2021-08-13). DOI: 10.11922/sciencedb.j00001.00273.

A dataset of district/county-level population distribution of China's six national censuses

WU Haoran¹, GAO Liang^{1*}, SONG Dongdong¹, YANG Yitao¹,

XU Changxing¹, YANG Xiaobao¹

1. Key Laboratory of Transport Industry of Big Data Application Technologies for Comprehensive Transport, Beijing Jiaotong University, Beijing 100044, P.R. China

*Email: lianggao@bjtu.edu.cn

Abstract: Spatialized population data are of great importance for supporting the studies on transportation, geography, socio-economy, sustainable development etc. In this paper, we collected the district/county-level population data of China's six population censuses starting from 1949. With the administrative division in 2015 (the year of the Sixth Population Census) as the benchmark of the geographical region of the districts and counties, we calibrated and spatialized the population data to obtain a dataset of district/county-level population distribution of China's six national censuses.

Keywords: population census; spatialization; district/county; China

Dataset Profile

Title	A dataset of district/county-level population distribution of China's six national censuses
Data corresponding author	GAO Liang (lianggao@bjtu.edu.cn)
Data authors	WU Haoran, GAO Liang, SONG Dongdong, YANG Yitao, XU Changxing, YANG Xiaobao
Time range	1953, 1964, 1982, 1990, 2000, 2010
Geographical scope	Mainland China
Data volume	67.6 MB
Data format	(*shp, *.cpg, *.dbf, *.prj, *.sbn, *.sbx, *.shx)
Data service system	< http://www.doi.org/10.11922/sciencedb.j00001.00273 >

Sources of funding	National Natural Science Foundation of China (71571017, 91646124, 71621001, 91746201)
Dataset composition	The dataset consists of one file, namely “county/district_ POP_ 20210602.rar”, covering the coded and attributed links to form a district/county level GIS population database of 31 provinces (municipalities, autonomous regions) (excluding Hong Kong, Macao, and Taiwan).