AustralianPoliticians: A new dataset of all Australian federal, state, and colonial politicians [*]

Rohan Alexander   *University of Toronto*

Understanding who is, and has been, in parliament is a critical first step before further analysis can occur. Nonetheless, there is no publicly available comprehensive dataset of those who have been members of Australia's federal, state, and colonial parliaments. In this paper we introduce a new dataset - AustralianPoliticians - which contains daily data on the members of parliament for the Australian Federal Parliament, the six state parliaments, the two territory parliaments, and the preceeding colonial parliaments. We discuss the creation of this dataset, and then show its validity through various existing datasets and internal consistency tests. Finally, we discuss some extensions to this dataset. This dataset will be of use to those interested in quantitative analysis of Australian politics.

Keywords: Australian politics, dataset, comparative politics

Introduction

A critical first step for modern quantitative political analysis is a publicly available comprehensive dataset. In this paper we introduce one such dataset - AustralianPoliticians - which documents everyone who has been a member of any Australian federal, state, and/or colonial parliament. By bringing together various sources in this way, researchers can use this dataset to answer critical questions related to representation. Researchers can also build extensions to this dataset that focus on their particular interest.

AustralianPoliticians is a record of everyone who has been a member of the Australian Federal Parliament (starting from 1901), the six state parliaments as well as their colonial antecedents, and the two territories. There are X unique people in the dataset, and the core dataset - `all` - provides basic biographical information about them. Related datasets are linked to this core dataset by a `uniqueID`, and deal with party, division, ministries, and finally translate the uniqueID into various other IDs used by other sources.

The datasets are up-to-date as of 1 July 2020, and will be maintained and updated on a monthly basis, by the author for the next five years. The latest version of the datasets are publicly available via GitHub, with an annual version lodged with the Australian Data Archive.

These datasets represent a critical contribution to Australian

---

Summary statistics

Creation and sources

Validity

Internal consistency

Existing datasets

Stylized example

Extensions