# Reviewer I

1. Aktuell verfolgt das Manuskript, wie oben beschrieben, zwei Ziele, einen substanziellen Beitrag und einen Datenbeitrag. Für eine Research Note ist dies meiner Meinung nach zu viel und wird hier dadurch deutlich, dass beide Beiträge in ihrer Umsetzung eher oberflächlich bleiben und Schwachstellen aufweisen. Die Autoren sollten sich daher entscheiden, ob sie einen substanziellen Beitrag zur Deskriptiven Repräsentation liefern, oder ob sie vollumfänglich auf den Datensatz eingehen möchten. Ich würde letzteres favorisieren und den Titel des Manuskripts entsprechend anpassen. Je nachdem, wofür sich die Autoren entscheiden, sind für die Überarbeitung des Manuskripts unterschiedliche der im Folgenden genannten Punkte relevant.

* Dazu Linhart: *Aus meiner Sicht ist eindeutig, dass das Vorstellen des Datensatzes im Vordergrund steht und die Auswertungen lediglich exemplarisch zeigen, welche Möglichkeiten der Datensatz bietet. Ich würde dennoch nicht auf die Anwendungen gänzlich verzichten - diese Art der Research Note hat sich in ähnlicher Form in der PVS bewährt. Vielleicht könnten Sie an der einen oder anderen Stelle im Manusript aber klarer machen, dass die Präsentation des Datensatzes im Vordergrund steht.*

1. Zur Rechtfertigung der Relevanz deskriptiver Repräsentation bedient sich das Manuskript argumentativ an verschiedenen Stellen besonders interessanter Merkmale, wie etwa dem Bildungsgrad, dem Migrationshintergrund, dem Einkommen oder auch dem Beruf von Abgeordneten. Siehe hierzu die Fragen im ersten Absatz der Einleitung oder verschiedene Teile des Theorieabschnitts. Diese Argumentationsweise weckt Erwartungen, die der Datensatz und die Analyse letztlich nicht erfüllen. Dies liegt vor allem daran, dass Bildung, Einkommen, Beruf, Wohnort (Stadt/Land) und Migrationsgeschichte gar nicht im Datensatz erfasst und viele der zur Analyse deskriptiver Repräsentation relevanter Fragen mit diesem folglich gar nicht zu beantworten sind. Außerdem wird auch der Fokus der deskriptiven Repräsentation dadurch motiviert, dass diese besonders dann für substanzielle Repräsentation bedeutsam ist, „wenn Gruppen eine distinkte gesellschaftliche Position und damit verbundene Alltagserfahrungen und politische Präferenzen aufweisen". Darauf bezogen werden die Merkmale Geschlecht, Alter und Herkunft zwar thematisiert. Gleichzeitig stechen eben diese hier eher als Nebenschauplätze heraus, vor allem da bei der Herkunft nur ein Teil der Migrationshistorie, lediglich der Geburtsort, erfasst wird, nicht aber die Geburtsorte der Eltern.

* Christian

1. Aus dem vorigen Punkt folgend, bleibt die Analyse der gewählten Merkmale, Geschlecht, Alter und Herkunft, dann doch eher blass. Insbesondere erscheinen die einfachen Zeitreihen zum Frauenanteil (Abbildung 2), zum Durchschnittsalter (Abbildung 4) sowie zum Geburtsort (Abbildung 5) wie übliche Verdächtige und wirken tendenziell eher unkreativ. Hier würde ich mir etwas ausgefallenere Anwendungen, ähnlich der Abbildung 3 wünschen, das erscheint mit wirklich neu und spannend. Warum nicht z.B. die im Fazit genannte Frage aufgreifen: „Gelingt es Männern, bei sonst gleichen Qualifikationen und Eigenschaften (z.B. Seniorität) auch in der Landespolitik eher als Frauen Spitzenpositionen zu besetzen?". Diese Frage sollte mit den vorhandenen Daten doch bearbeitbar sein.

* Nochmal mit Jochen & Christian besprechen

1. Das zweite, substanzielle Ziel des Manuskripts wird m.E. auch deshalb nur unzureichend erfüllt, da zwar deskriptive Befunde anhand der neuen Datenlage gezeigt und beschrieben werden. Diese werden aber kaum bis gar nicht vor dem Hintergrund der zuvor umfangreich beschriebenen Relevanz der deskriptiven für die substanzielle Repräsentation diskutiert. Hier klafft eine Lücke zwischen dem Theorieteil und der Analyse. Was lernen wir aus den Befunden und welche Schlüsse können wir daraus insbesondere mit Blick auf substanzielle Repräsentation ziehen?

* Hier können wir einen Aufschlag machen!

1. Mit Blick auf die oben genannten, wünschenswerten, aber fehlenden Merkmale kommt die Frage auf, warum diese nicht im Datensatz vorhanden sind? Die Datenlage ist wie sie ist und die im Datensatz vorhandenen Information sind nichtsdestotrotz von großem Nutzen. Dennoch, um diese Frage überzeugend zu beantworten und Verständnis zu generieren, bedarf es mehr Informationen zur Datenerhebung und -verfügbarkeit. Aus welchen Quellen wurden die Daten genau zusammengeführt? Wurden Informationen aus Wikipedia-Einträgen manuell oder automatisiert extrahiert? Welche Teile der Einträge wurden verwendet, die gesamten Seiten, die einleitenden Absätze, die Infoboxen? Wurde nur auf Wikipedia oder auch Wikidata zurückgegriffen? Welche Veröffentlichungen der Landesparlamente und daraus welche Teile wurden genutzt? Wie wurden diese Informationen beschafft und digital zugänglich gemacht? Gab es Kodierer, entsprechende Codebücher, und wie steht es um die Intercoderreliabilität? In welchem Format werden die Daten zur Verfügung gestellt? Die Einführung eines Datensatzes sollte Fragen dieser Art besprechen, auch wenn sich Antworten darauf teilweise im Anhang unterbringen lassen.

* Aus welchen Quellen wurden die Daten genau zusammengeführt?
* Wurden Informationen aus Wikipedia-Einträgen manuell oder automatisiert extrahiert?
* Welche Teile der Einträge wurden verwendet, die gesamten Seiten, die einleitenden Absätze, die Infoboxen? Wurde nur auf Wikipedia oder auch Wikidata zurückgegriffen?
* Welche Veröffentlichungen der Landesparlamente und daraus welche Teile wurden genutzt?
* Wie wurden diese Informationen beschafft und digital zugänglich gemacht?
* Gab es Kodierer, entsprechende Codebücher, und wie steht es um die Intercoderreliabilität? In welchem Format werden die Daten zur Verfügung gestellt?

1. Ich sehe den Fokus auf die Landes- anstelle der Bundesebene als einen zentralen Beitrag dieses Datensatzes. Gleichzeitig wird im Manuskript bislang eher wenig dazu gesagt, warum Daten auf Länderebene besonders relevant sind und welchen Nutzen sie im Vergleich zu den in Hülle und Fülle vorliegenden Daten auf Bundesebene liefern. Hier bietet sich ggf. ein etwas detaillierterer Vergleich zur Bundesebene an, ggf. im Analyseabschnitt.

* Jochen und Christian mal fragen, was hier am besten zu nennen ist?

1. Neben dem Fokus auf Länderebene liegt der eigentliche Datenschatz für mich in den Informationen zu Mandat, Ämtern und Funktionen. Dieser Beitrag wird an mehreren Stellen auch explizit hervorgehoben, z.B. im ersten Absatz des Fazits. Abgesehen vom Datensatz von Turner-Zwinkels et al. (2021), ist mir kein Datensatz bekannt, der überhaupt derartige Informationen beinhaltet. Hier würde sich ggf. auch ein direkter Vergleich zu Turner-Zwinkels et al. (2021) lohnen, um den Beitrag noch stärker herauszustellen. Statt auf die üblichen und oft thematisierten Merkmale Geschlecht, Alter und Herkunft im Kontext deskriptiver Repräsentation einzugehen, würde ich den Artikel viel stärker auf diesen Teil des Datensatzes münzen. Wie oben beschrieben, geschieht das mit Abbildung 3 ja bereits in Teilen.

Hier außerdem noch ein paar kleinere Anmerkungen:

1. Bei der Beschreibung der Datenstruktur ist zunächst die Rede von einer Liste aller Personen. In Abbildung 1 erscheinen dann aber zwei Listen. Dies entspricht auch nicht dem relationalen Standard einer Datenbank. Stattdessen sollten die beiden Listen zusammengefügt und die Kabinettsinformationen, ähnlich den Präsidiums- und Ausschussdaten, gesondert aufgeführt und lediglich via „id" zugeführt werden. Ansonsten ist bspw. das Feld „name" in zwei Datensätzen verfügbar.
   * @Daniel, sollen wir das übernehmen? Sonst begründen warum nicht.
2. Göbel/Munzert 2020 werden mit dem R Paket zitiert. Hier sollte stattdessen der entsprechende Zeitschriftenaufsatz zitiert werden: Göbel, S., & Munzert, S. (2022). The Comparative Legislators Database. British Journal of Political Science, 52(3), 1398-1408. doi:10.1017/S0007123420000897
   * Korrigieren
3. Die Autoren schildern, dass sie unter anderem das R Paket „genderizeR" verwendet haben. Daraus lässt sich schließen, dass das Geschlecht der Abgeordneten anhand des Namens geschätzt wurde. Vor diesem Hintergrund wird die oben erwähnte genauere Beschreibung der Datenerhebung sowie Informationen hinsichtlich der Validierung noch zentraler. Informationen zu den R Paketen kann darüber hinaus auch im Anhang verschwinden.
   * Gender auf Grundlage von Namen geschätzt

Reviewer #2:

1. Das Potenzial des Datensatzes wird in der deskriptiven Analyse nicht klar herausgearbeitet. Immer wieder wird im Artikel überzeugend das Argument präsentiert, der Datensatz biete tagesaktuelle Daten zu den Ämtern, die Landtagsmitglieder übernehmen. In der deskriptiven Analyse spielt diese Komponente jedoch überhaupt keine Rolle. Es werden bspw. bekannte Fakten zur Repräsentation von Frauen in den Landtagen im Zeitverlauf aus früheren Studien wiederholt. Oder die Zeiteffekte werden als nicht informative Variation abgetan (bei der Diskussion des zackigen Verlaufs des Durchschnittsalters, S. 10). Hilfreich wäre, wenn der Fokus der deskriptiven Analyse so gewählt wäre, dass der Mehrwert des Datensatzes ersichtlich ist. Alternativ wäre es auch möglich, in einem zusätzlichen Abschnitt das Potenzial für zukünftige Forschung genauer auszuführen. Aktuell finden sich Überlegungen zur Verwendung der originellen Komponente des Datensatzes nur in zwei Sätzen im Fazit. Insgesamt fehlt eine überzeugende Darlegung des Mehrwerts der tagesgenauen Daten.
2. Um die Entstehung des Datensatzes transparent zu machen, fehlen noch einige Informationen. Insbesondere kamen für mich folgende Fragen auf: (1) Die Daten beginnen 1990 - doch wann enden sie im aktuellen Stand? (2) Welche Datenquellen wurden verwendet bzw. wo können die Datenquellen später eingesehen werden, sofern eine Darstellung im Rahmen der Research Note nicht möglich ist? (3) Welche Schritte der Validierung der Daten wurden durchgeführt (angesprochen auf S. 5, aber nicht erläutert)?
   * Start und Enddaten expliziter machen
   * Welche Datenquellen? Hierarchie darstellen. Auf Dokumentation verweisen
   * Validierung

Kleinere Anregungen und Anmerkungen:

(3) Mit der Publikation von Höhmann 2023 in European Political Science Review („Political parties, issue salience, and the appointment of women cabinet members") existiert bereits ein Datensatz sowie eine Auswertung zum Geschlecht von Ministerinnen in den Bundesländern. Dies könnte in die Liste der existierenden Quellen noch aufgenommen werden.

* aufnehmen

(4) Um eine breitere Leserinnenschaft anzusprechen, wäre es hilfreich, über den deutschen Fokus hinauszugehen und darzulegen, welche Relevanz die vorliegenden Daten auch für die international vergleichende Forschung haben könnten.

* Wie oben: Länderparlamente & Generalisierbarkeit klären

(5) Für die Auswertung des Durchschnittsalters wäre als Referenzwert eine Information über das Durchschnittsalter der Bevölkerung hilfreich, um zu erkennen, ob die vorliegenden Trends nur der Alterung der Gesellschaft entsprechen oder ob Landtagsabgeordnete systematisch älter werden (wie auf S. 10 oben angedeutet).

- einfügen

(6) Ich würde gerne eine kurze Diskussion lesen, inwiefern aus dem Geburtsort in West-/Ostdeutschland eine Zuordnung als west-/ostdeutsche Person erfolgen kann im Sinne einer deskriptiven Repräsentationsfähigkeit. Mensch könnte argumentieren, dass ein Stichtag zum 18. Geburtstag sinnvoller wäre; oder aber, dass auch beide Elternteile im jeweiligen Landesteil aufgewachsen sein sollten. Dann wäre klar, dass eine abgeschlossene Sozialisierung im entsprechenden politischen und gesellschaftlichen Raum stattgefunden hat. Die Autorinnen sollten ihre Operationalisierung begründen.

- Geburtsort in Ostdeutschland = Ostsozialisierung?