Who are the loneliest Americans?

Draft

Joe Marlo

2020-09-16

Abstract

Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum. Arcu dictum varius duis at consectetur lorem donec massa. Vestibulum lectus mauris ultrices eros in cursus turpis massa tincidunt. Dui accumsan sit amet nulla. At tellus at urna condimentum mattis pellentesque id nibh tortor. Interdum velit euismod in pellentesque massa placerat. Massa sed elementum tempus egestas. Turpis egestas integer eget aliquet nibh. Cursus eget nunc scelerisque viverra mauris in aliquam sem. Aliquet bibendum enim facilisis gravida neque. Sagittis eu volutpat odio facilisis mauris sit amet massa. Aliquet lectus proin nibh nisl condimentum. A cras semper auctor neque vitae. Lobortis scelerisque fermentum dui faucibus in ornare quam viverra orci. Est lorem ipsum dolor sit amet consectetur adipiscing elit.

Contents

Background 2

Literature 2

Methodology 4

Distance measures 4

Clustering 5

Modeling 10

Associated demographics 15

Conclusion 17

Appendix 18

Background

Time spent alone has been increasing among Americans. This can have numerous health effects (x,y,z) and it may be impacting subpopulations differently. Data from the American Time Use Survey shows the mean amount of time spent on non-work activities with no other person present has steadily increased from ~295min per day to ~330min per day from 2003 to 2018/

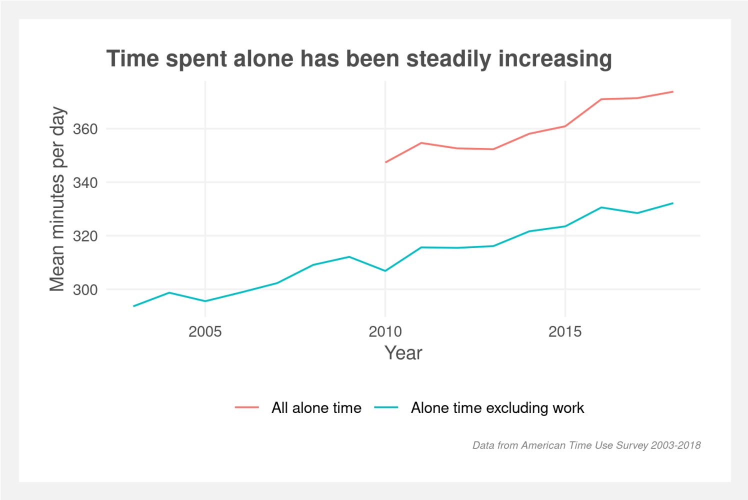

Figure 1: Mean alone time 2003-2018

It's intuitive that this increase may not be evenly distributed across the population. Rather than subdividing the population by demographics, the population can be divided using sequence analysis and unsupervised learning techniques to find clusters of similar time-use patterns. These in turn may represent distinct demographic groups (e.g. a cluster with large amounts of time spent on education consists mostly of sub 25 year olds) but are not direct measurements of demographics. This clustering methodology allows demarcation of groups based on their activity and may capture groups such as students, workers, and the elderly.

Research question

> Are increases in time spent alone equally affecting different subpopulations of Americans?

Literature

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Mattis enim ut tellus elementum sagittis vitae et. Odio pellentesque diam volutpat commodo sed egestas egestas fringilla. Consectetur lorem donec massa sapien faucibus. Euismod nisi porta lorem mollis aliquam ut porttitor leo a. Metus dictum at tempor commodo. Aliquam etiam erat velit scelerisque in dictum non consectetur. Volutpat odio facilisis mauris sit amet massa vitae tortor condimentum. Nibh venenatis cras sed felis eget velit.

Facilisi cras fermentum odio eu. Quis enim lobortis scelerisque fermentum dui faucibus in. Commodo elit at imperdiet dui. Mi quis hendrerit dolor magna eget. Scelerisque purus semper eget duis at tellus at. Aliquet bibendum enim facilisis gravida neque. Sagittis eu volutpat odio facilisis mauris sit amet massa. Aliquet lectus proin nibh nisl condimentum. A cras semper auctor neque vitae. Lobortis scelerisque fermentum dui faucibus in ornare quam viverra orci. Est lorem ipsum dolor sit amet consectetur adipiscing elit.

Arcu dictum varius duis at consectetur lorem donec massa. Vestibulum lectus mauris ultrices eros in cursus turpis massa tincidunt. Dui accumsan sit amet nulla. At tellus at urna condimentum mattis pellentesque id nibh tortor. Interdum velit euismod in pellentesque massa placerat. Massa sed elementum tempus egestas. Turpis egestas integer eget aliquet nibh. Cursus eget nunc scelerisque viverra mauris in aliquam sem.

Defining loneliness: - https://www.researchgate.net/profile/Daniel_Russell4/publication/271766646_The_measurement_of_loneliness/links/56bbc30408ae7be8798be595.pdf
- "some suggest the evidence of social decline is based on flawed indicators" https://doi.apa.org/doiLanding?doi=10.1037%2F1040-3590.7.3.286

Trends: - loneliness over time meta-analysis https://journals.sagepub.com/doi/full/10.1177/0146167214557007?casa_token=qprWkQrW1p4AAAAA%3AtWSkTxcYjZJPx_4y5vtcKLKV6kJQQbOQ4lHz9T_7UKYya5_QN1cD12cxv6Z_JHL7fcf-sG1vrfor
- "research that demonstrates declining social engagement; in comparison with decades past, people are less likely to join clubs, have fewer confidants, and are less likely to perceive others as trustworthy" https://journals.sagepub.com/doi/10.1177/000312240607100301
- "Social engagement through the internet, however, could be replacing traditional forms of sociability" https://journals.sagepub.com/doi/10.1177/1948550612469233
- https://books.google.com/books?hl=en&lr=&id=ySBrAAAAMAAJ&oi=fnd&pg=PA13&dq=loneliness+american&ots=dldWStCMS6&sig=NJGTmhoDtg7uXdLMX7pTA3oCgIc#v=onepage&q=loneliness%20american&f=false

Elderly: - "whether loneliness increases in old age, and if so, whether it relates to ageing itself, to time trends or to cohort effects and ..." https://academic.oup.com/ageing/article/28/5/491/36243
- "loneliness explained the excess risk of depression in the widowed." https://onlinelibrary.wiley.com/doi/abs/10.1002/gps.2181?casa_token=0ym35WpAkZ0AAAAA%3ABdmKXAZTQ6C5xKe538Jvr9PPypZem0_zhZYvgOz7W6fKbK2Cklqw_Xslq8sQk5xF7sPCRAqyeVu35w
- health outcomes https://journals.sagepub.com/doi/abs/10.1177/0898264305280993?casa_token=HF41nPUAZqUAAAAA:8zg6wBtozRz52BPPv6m-pkxQ1QlGCDJfxjmY_Uw63p9C7AsQDPT6C0MpFGcpasSnqqGgv4FRsBIo
- https://journals.sagepub.com/doi/pdf/10.1177/0164027587094001
- causal https://academic.oup.com/geronj/article-abstract/40/4/487/593670

Pop: - https://www.theatlantic.com/magazine/archive/2012/05/is-facebook-making-us-lonely/308930/
- https://www.nytimes.com/2006/07/02/weekinreview/02fountain.html
- A PLAGUE OF LONELINESS: Americans were already lonely. Then COVID-19 hit. By: DUCHARME, JAMIE, Time International (South Pacific Edition), 08180628, 6/22/2020, Vol. 195, Issue 23/24

Methods - Abbot: https://www.jstor.org/stable/2780695
- https://www.sciencedirect.com/science/article/pii/S0965856416303470

Alone time definition

The American Time Use Survey tracks alone time via a computation of other collected variables. For each activity — except those noted below — the BLS tracks the number of participants present during the activity. Alone time is tallied only during activities for which only the primary respondent is physically present.

The benefit of this approach is detailed data on the length of time and description of the activity in which the person is alone. The shortcomings of this is approach is that it only pertains to physicality. Therefore activities such as phone or video calls will be labeled as 'alone' unless there is an additionally person physically present. Additionally, a few activities are specifically excluded from the tally including:

- Working
- Sleeping
- Washing, dressing, or grooming
- Personal/private activities
- Any time in which the respondent refused to provide activity detail

The question posed to respondents to define who was present:
> "Who was in the room with you / Who accompanied you"

The BLS also includes another variable, `TRTALONE_WK` which is similar to `TRTALONE` but includes alone time during work activities. This is excluded for the analysis as it is only available 2010-2018.

Methodology

Clustering methods will be used to determine similar sequences of how individuals spend their day irrespective of alone time. Primary techniques will be using string editing techniques ((optimal matching), Abbot) and secondary techniques are model-based clustering…

String editing techniques will start by aggregating the different types of activities from 465 specific activities into 15 activities based on their hierarchical definitions provided by the BLS. These 15 activities will then be recoded as single character strings representing how an individual spends each 30 minute period of their day. Their full day's activity is represented by the resulting 48 character string.

…model based…

Data detail

Data comes from The American Time Use which surveys how Americans spend their time. The diary (`atusact_0318` file) and CPS (`atuscps_0318` file) data are used from the 2003-2018 Multi-Year Interview dataset. The data details each minute of the respondent's day by mapping it to a list of 465 activities. The author then aggregated these 465 activities into 15 activities based on the Bureau of Labor Statistics' (BLS) hierarchical definitions and the author's judgment. See Appendix Table 3 for the aggregation mapping. Additionally, to reduce computation load, each respondents' day was summarized into 48 thirty-minute windows representing the modal activity during the window.

Time use varies greatly between week and weekend days so only weekdays are included in the analysis. Similarly, holidays are excluded.

Distance measures

String editing techniques will start by aggregating the different types of activities from 465 specific activities into 15 activities based on their hierarchical definitions provided by the BLS. These 15 activities will then be recoded as single character strings representing how an individual spends each 30 minute period of their day. Their full day's activity is represented by the resulting 48 character string.

The distance between the respondents string sequences can be calculated using a number of different string distance measures:

- Levenshtein distance: insertions, deletions or substitutions
- Restricted Damerau-Levenshtein distance (OSA): insertions, deletions or substitutions of a single character, or transposition of two adjacent characters
- Hamming distanc: substitutions only
- Longest common subsequence: insertions and deletions only

Each of these have advantages and disadvantages that will be explored along with their impact on the final clustering.

...expand on implications of each distance measure...

Clustering

Discuss clustering methods. Facilisi cras fermentum odio eu. Quis enim lobortis scelerisque fermentum dui faucibus in. Commodo elit at imperdiet dui. Mi quis hendrerit dolor magna eget. Scelerisque purus semper eget duis at tellus at. Aliquet bibendum enim facilisis gravida neque. Sagittis eu volutpat odio facilisis mauris sit amet massa. Aliquet lectus proin nibh nisl condimentum. A cras semper auctor neque vitae. Lobortis scelerisque fermentum dui faucibus in ornare quam viverra orci. Est lorem ipsum dolor sit amet consectetur adipiscing elit.

Efficacy of cross-sectional clustering

The data consists of cross-sectional observations of individual's time use. The clusters are computed across years. Therefore, no single respondent represents more than one year but individual clusters consists of multiple years.

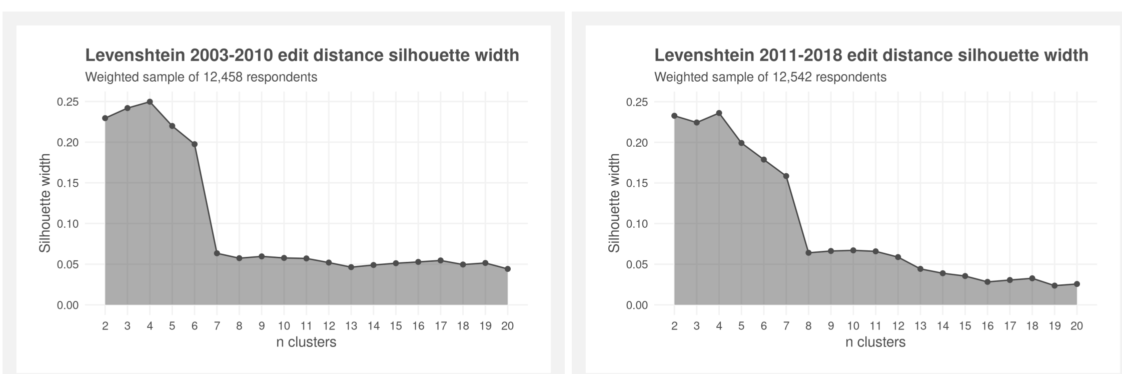

Figure 2: Silhouette comparison

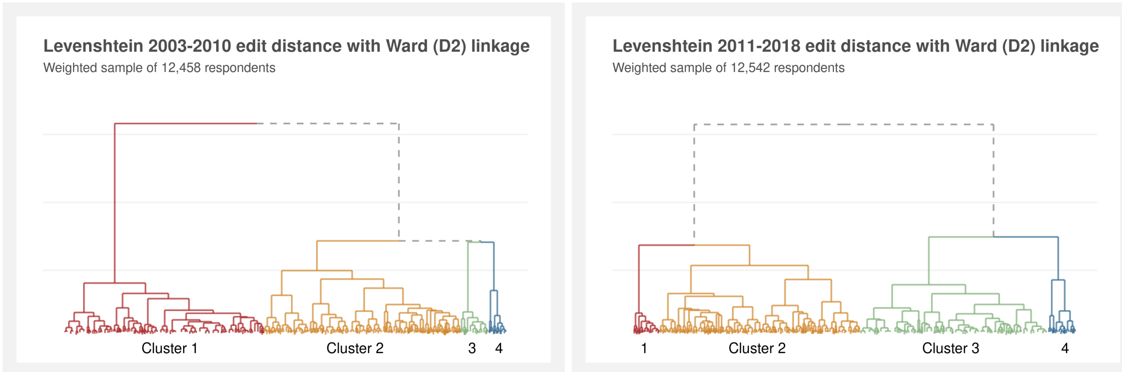

Figure 3: Dendrogram comparison

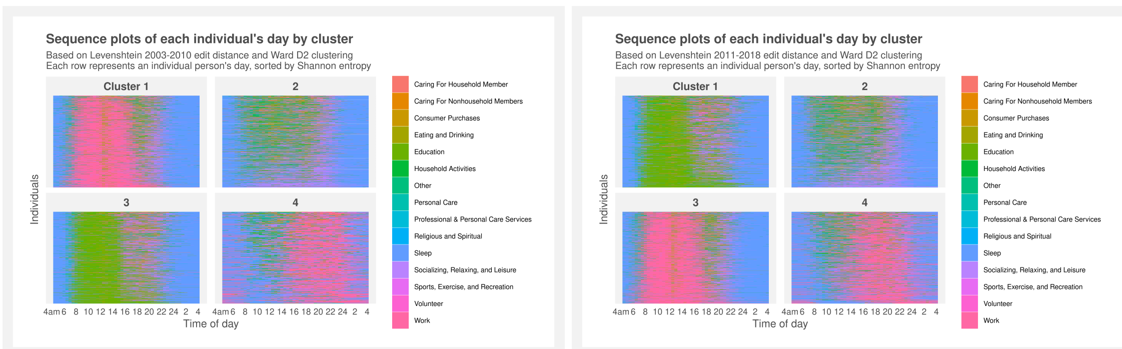

Figure 4: Sequence plots comparison

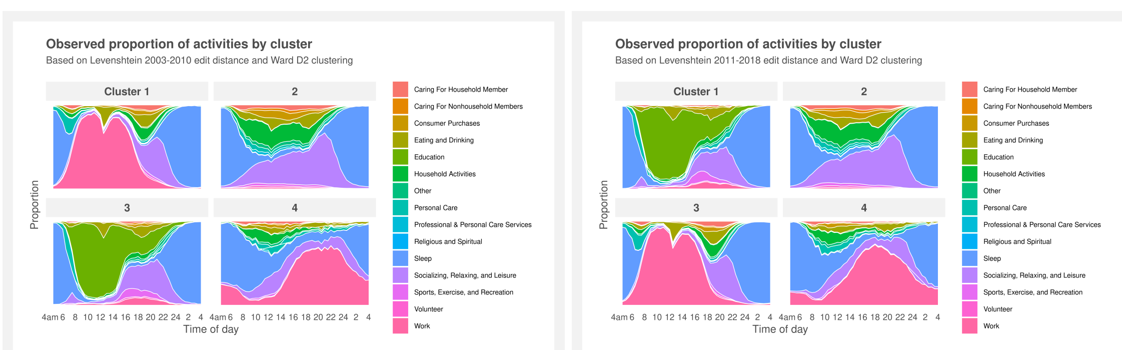

Figure 5: Proportion plots comparison

Final clusters

The final clusters were created using a weighted sample of 25,000 respondents from the 2003-2018 surveys.

Discuss silhouettes and dendrogram. Metus dictum at tempor commodo. Aliquam etiam erat velit scelerisque in dictum non consectetur. Volutpat odio facilisis mauris sit amet massa vitae tortor condimentum. Nibh venenatis cras sed felis eget velit.

Volutpat odio facilisis mauris sit amet massa vitae tortor condimentum. Nibh venenatis cras sed felis eget velit. Mattis enim ut tellus elementum sagittis vitae et. Odio pellentesque diam volutpat commodo sed egestas egestas fringilla. Consectetur lorem donec massa sapien faucibus. Euismod nisi porta lorem mollis aliquam ut porttitor leo a. Metus dictum at tempor commodo. Aliquam etiam erat velit scelerisque in dictum non consectetur.

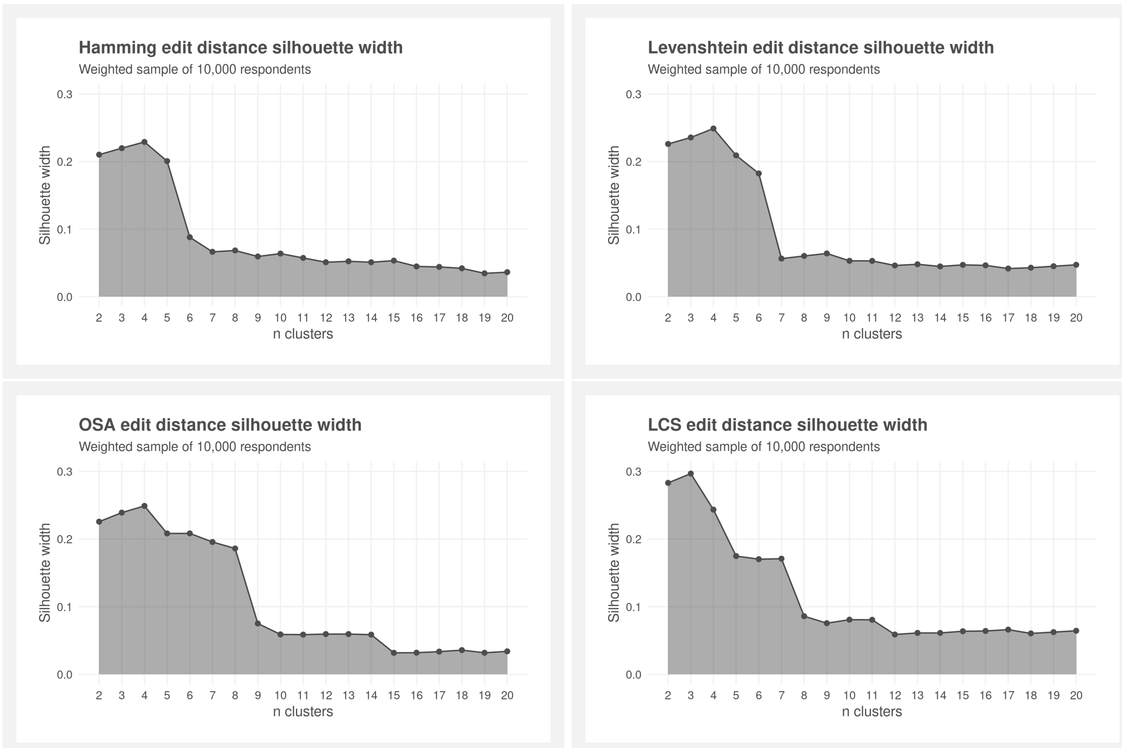

Figure 6: Silhouette distances

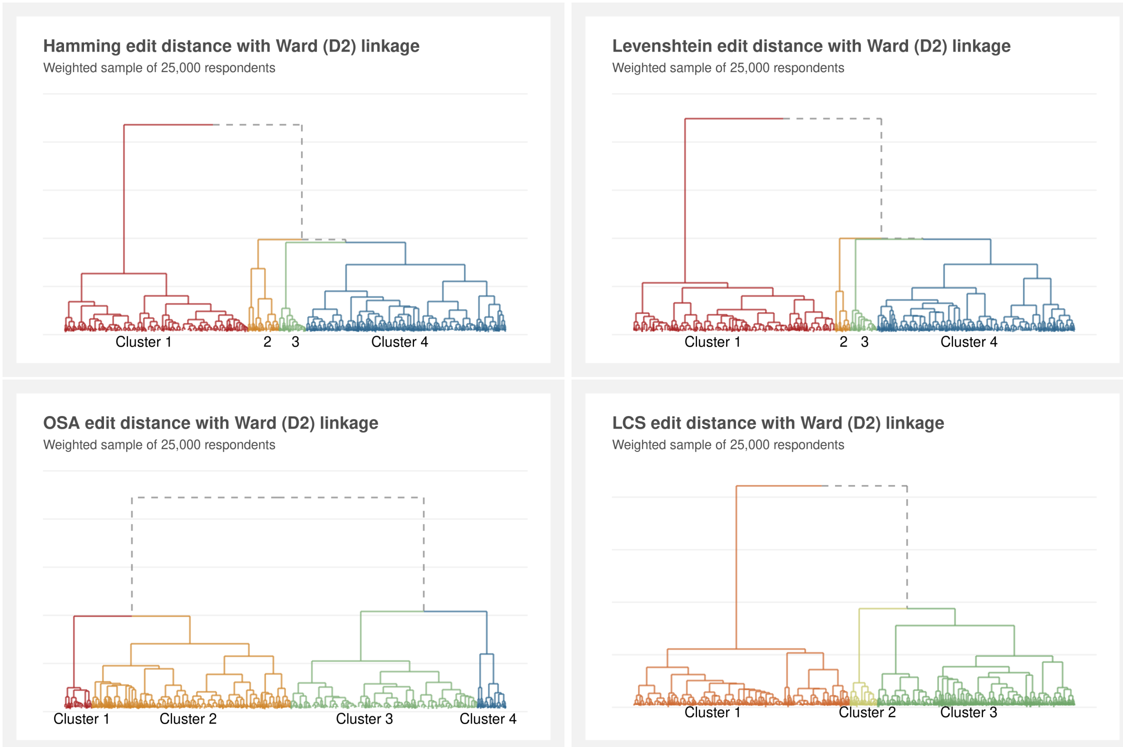

Figure 7: Dendrograms

Discuss sequence plots and patterns. Mattis enim ut tellus elementum sagittis vitae et. Odio pellentesque diam volutpat commodo sed egestas egestas fringilla. Consectetur lorem donec massa sapien faucibus. Euismod nisi porta lorem mollis aliquam ut porttitor leo a. Metus dictum at tempor commodo. Aliquam etiam erat velit scelerisque in dictum non consectetur. Volutpat odio facilisis mauris sit amet massa vitae tortor condimentum. Nibh venenatis cras sed felis eget velit.

- LCS materially different: making no distinct between daytime and nighttime workers

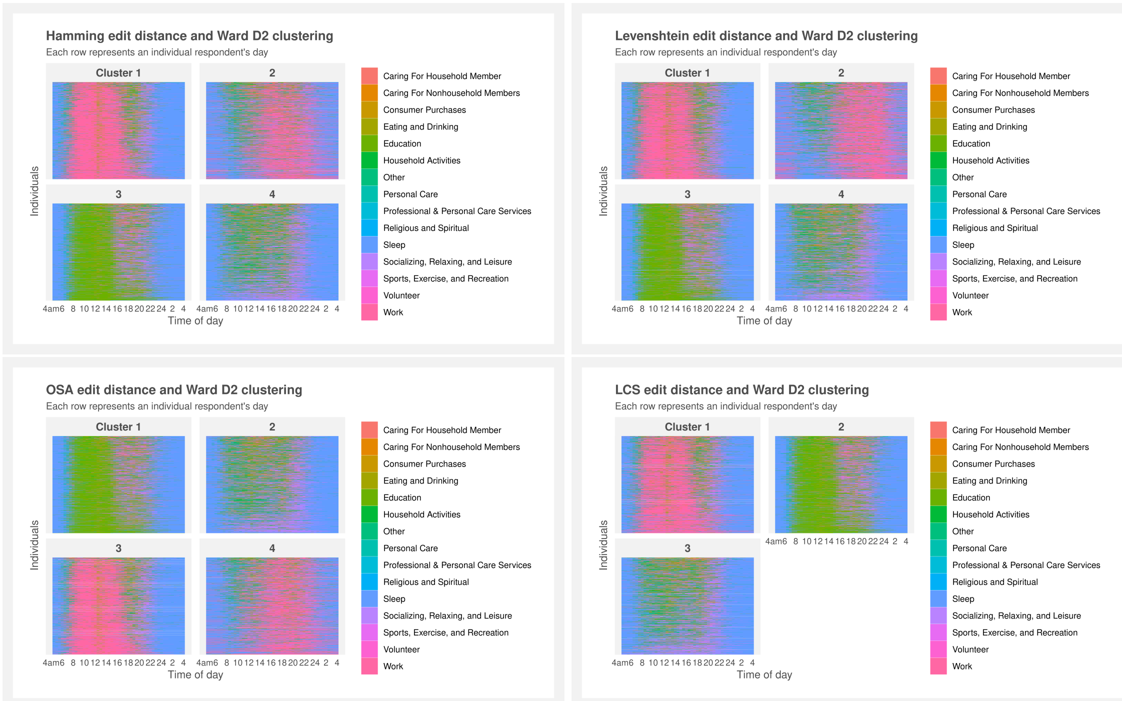

Figure 8: Sequence plots

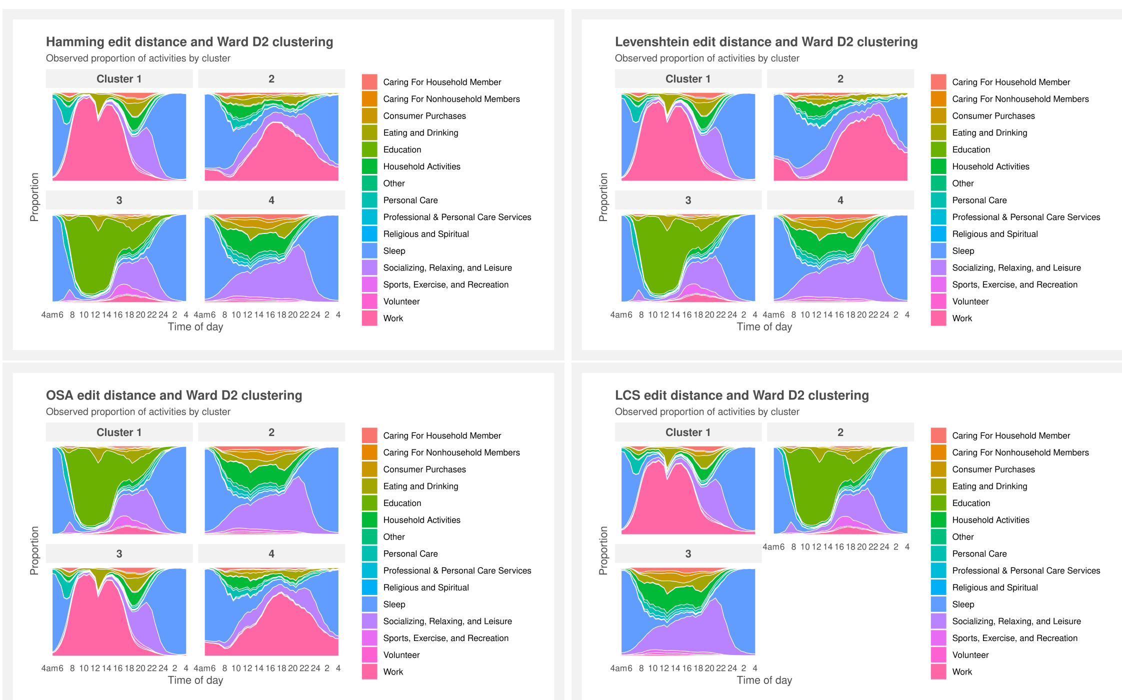

Figure 9: Proportion plots

Categorization of clusters

The sequence and proportion plots illustrate the contents of each cluster are mostly consistent across the edit distance methods with the exception of LCS. Each cluster can therefore be identified across methods using their common characteristics: one cluster consisting of mostly daytime work will now be labeled as “Day workers”, a second cluster of mostly evening and night work “Night workers”, a third dominated by education “Students”, and then the fourth contains a mixture of activities with no dominate characteristics so this will be referred to as “Uncategorized.”

Distance measure comparison

Visual inspections of the sequence and proportion plots appear to show consistency across the various edit distance measures. This is verified by examining the agreement in cluster membership. First, 90% of the respondents are clustered into the “same” cluster across methods. The remaining ∼10% match to two separate clusters, and a small amount match to three. This is most likely due to the LCS method optimizing for a three cluster solution.

The pattern is clearer examined on a respondent-level basis. The right plot in Figure # shows the cluster assignment per respondent per method. Homogenous colors horizontally indicate full agreement across the edit distance measures.

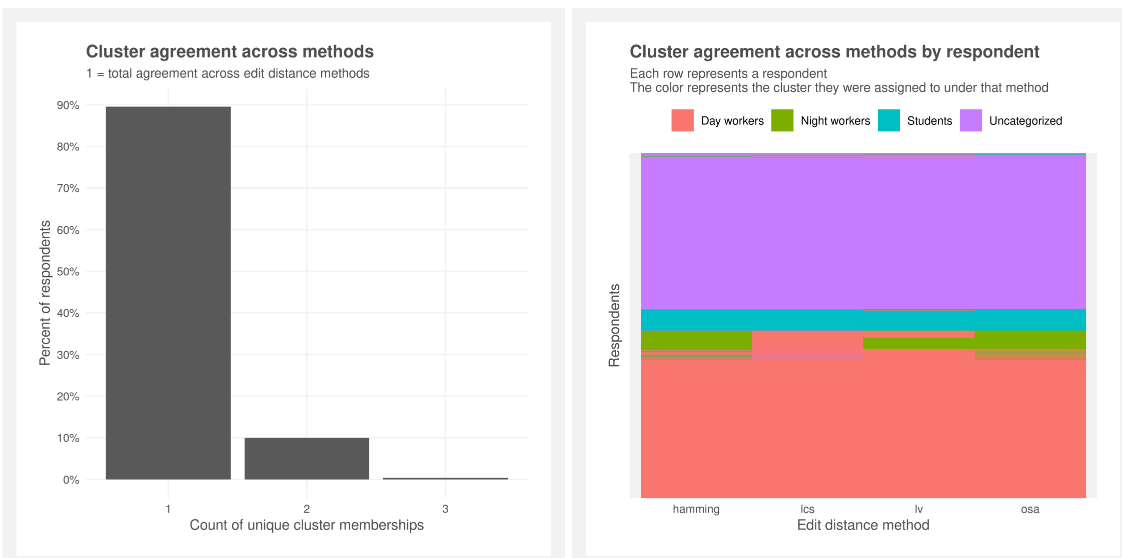

Figure 10: Cluster agreement across methods

Modeling

This is an exercise in comparing the implications of the edit distance methods. Each edit distance method will be modeled seperately, however the goal is also to understand how alone time has varied across different groups (i.e. clusters). Therefore, the unit of interest is the slope of alone time.

Single-level and multilevel models are both appropriate. Multilevel models “account for individual- and group-level variation in estimating group-level regression coefficients” and “estimate regression coefficients for particular groups” (Gelman Hill 2006). However, the clustering methods are signaling only four clusters are optimal. Gelman and Hill also note that “when the number of groups is small (less than five, say), there is typically not enough information to accurately estimate group-level variation. As a result, multilevel models in this setting typically gain little beyond classical varying-coefficient models.”

As such, the approach is to fit both single- and multilevel models and draw comparisons. This is multiplicative to comparing the edit distance methods so there will be 19 total models fit: 15 single-level models for each combination of edit distance and cluster membership (no pooling) plus four multilevel models, one for each edit distance method.

Count data and overdispersion

Describe data. egestas fringilla. Consectetur lorem donec massa sapien faucibus. Euismod nisi porta lorem mollis aliquam ut porttitor leo a. Metus dictum at tempor commodo. Aliquam etiam erat velit scelerisque in dictum non consectetur. Volutpat odio facilisis mauris sit amet massa vitae tortor condimentum. Nibh venenatis cras sed felis eget velit.

Not zero inflated. Only 4% of observations are zero

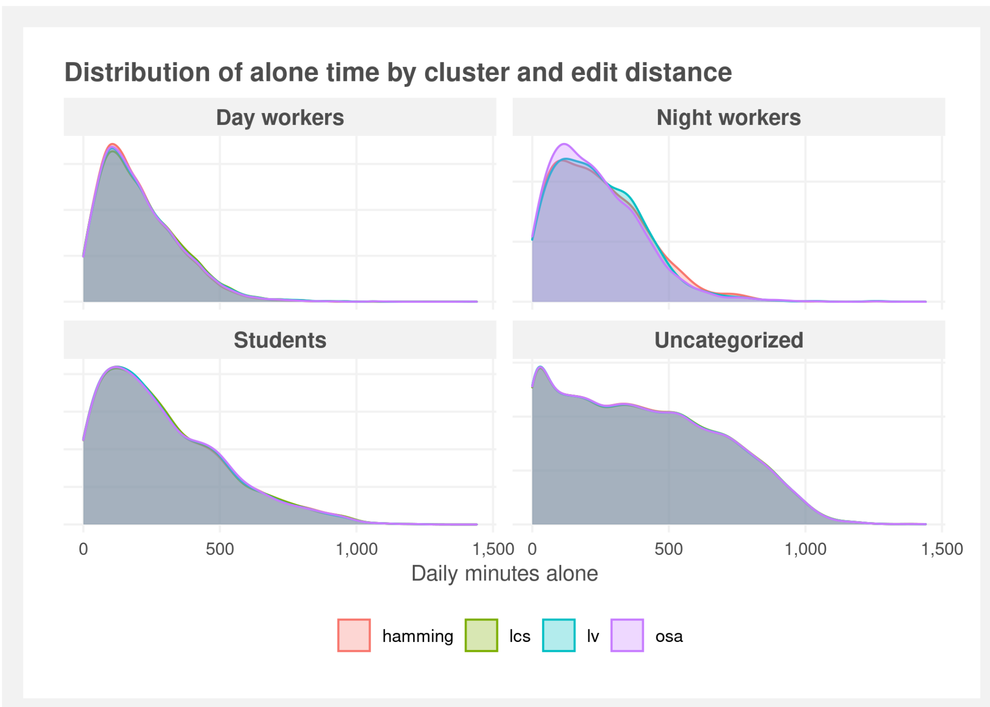

Figure 11: Densities of alone time by cluster

Overdispersion is an issue. The mean daily alone minutes range from 202 to 400 per each cluster and the variances range from 19,724 to 81,265. This violates an assumption of Poisson models: mean equals variance. It suggests that the standard errors in the Poisson model will be severely underestimated.

Dispersion can also be defined as estimated dispersion via Gelman and Hill (2006, pg 114):

$$dispersion\,ratio = \frac{1}{n-k}\sum_{i=1}^{n} z_i^2$$
$$n = \#\,of\ data\,points$$
$$k = \#\,of\ predictors$$
$$z = standardized\ residuals = \frac{y_i - y^i}{sd(y^i)}$$

Fitting a Poisson multilevel model to this data results in a dispersion ratio of 151.2 — far larger than 1 — indicating a quasi-Poisson or negative binomial model is necessary. Otherwise, the standard errors will need to be corrected by multiplying by a factor of $\sqrt{151} = 12.3$.

Multiple single-level quasi-Poissons

15 single-level quasi-Poissons models are fit to the data individually to estimate the effect of `Year` on per cluster per distance method. The models are fit stipulating the mean-variance relationship is $\sigma^2 = \delta\mu$.

Model equation:

$$log(\lambda) = \alpha + \beta_1 X_1 \qquad \lambda = mean\,count$$
$$X_1 = alone\ time$$
$$and \qquad \sigma^2 = variance$$
$$\delta = dispersion\,parameter$$
$$\sigma^2 = \delta\mu \qquad \mu = mean$$

The model form in R syntax is simply:

```
Alone time ~ year
```

Briefly discuss model fit. Aliquet bibendum enim facilisis gravida neque. Sagittis eu volutpat odio facilisis mauris sit amet massa. Aliquet lectus proin nibh nisl condimentum. A cras semper auctor neque vitae. Lobortis scelerisque fermentum dui faucibus in ornare quam viverra orci. Est lorem ipsum dolor sit amet consectetur adipiscing elit. The standard errors on thee coefficient are as large as the estimates themselves, indicating there is far too much noise in this signal to be certain of the pattern.

The coefficient for year is the difference in log counts. This can be also interpreted as the percentage increase in alone minutes per year. The only cluster to consistently and significantly differ from zero is the “Students” cluster. These point estimates range from -0.010 to -0.015, meaning the mean amount of alone time for students decreased at a rate of 1.0 to 1.5% per year from 2003 to 2018.

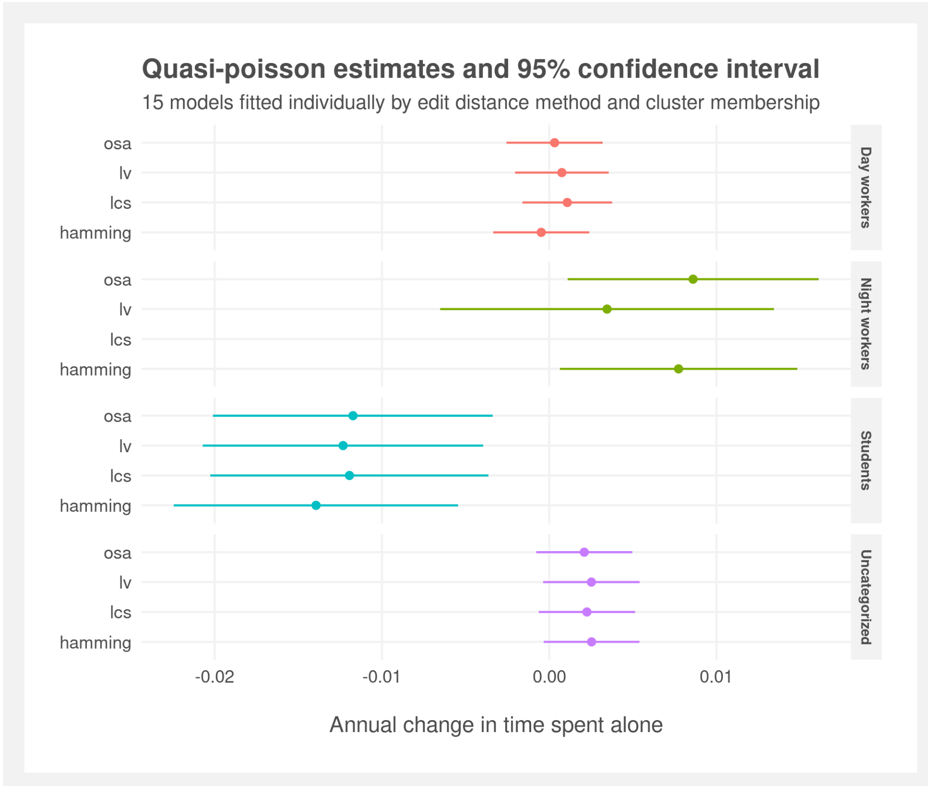

Figure 12: Quasi-Poisson estimates

Multilevel quasi-Poissons

Cluster as varying intercept and Year as fixed and random slope

varying intercepts and varying slope which themselves vary by cluster Correlated random intercept and slope

Model equation:

The model form in R syntax:

```
Alone time ~ cluster + (cluster | year )
```

Model fit

Describe diagnostics, residuals, etc. Facilisi cras fermentum odio eu. Quis enim lobortis scelerisque fermentum dui faucibus in. Commodo elit at imperdiet dui. Mi quis hendrerit dolor magna eget. Scelerisque purus semper eget duis at tellus at. Aliquet bibendum enim facilisis gravida neque. Sagittis eu volutpat odio facilisis mauris sit amet massa. Aliquet lectus proin nibh nisl condimentum. A cras semper auctor neque vitae. Lobortis scelerisque fermentum dui faucibus in ornare quam viverra orci. Est lorem ipsum dolor sit amet consectetur adipiscing elit.

Table 1: Multilevel quasi-Poisson model diagnostics

effect	group	term	estimate	std.error	df
fixed	fixed	(Intercept)	5.623	0.1261	24,995
fixed	fixed	year	-0.0008	0.0034	24,995
ran_pars	cluster	sd_(Intercept)	0.2509	NA	NA
ran_pars	cluster	cor_year.(Intercept)	-0.0105	NA	NA
ran_pars	cluster	sd_year	0.0062	NA	NA
ran_pars	Residual	sd_Observation	12.29	NA	NA

Model discussion

BLUPs are similar to the single-level models, however with some shrinkage towards zero. Discuss issues. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Discuss interpretation of effects. dio pellentesque diam volutpat commodo sed egestas egestas fringilla. Consectetur lorem donec massa sapien faucibus. Euismod nisi porta lorem mollis aliquam ut porttitor leo a. Metus dictum at tempor commodo. Aliquam etiam erat velit scelerisque in dictum non consectetur.

Discuss residuals

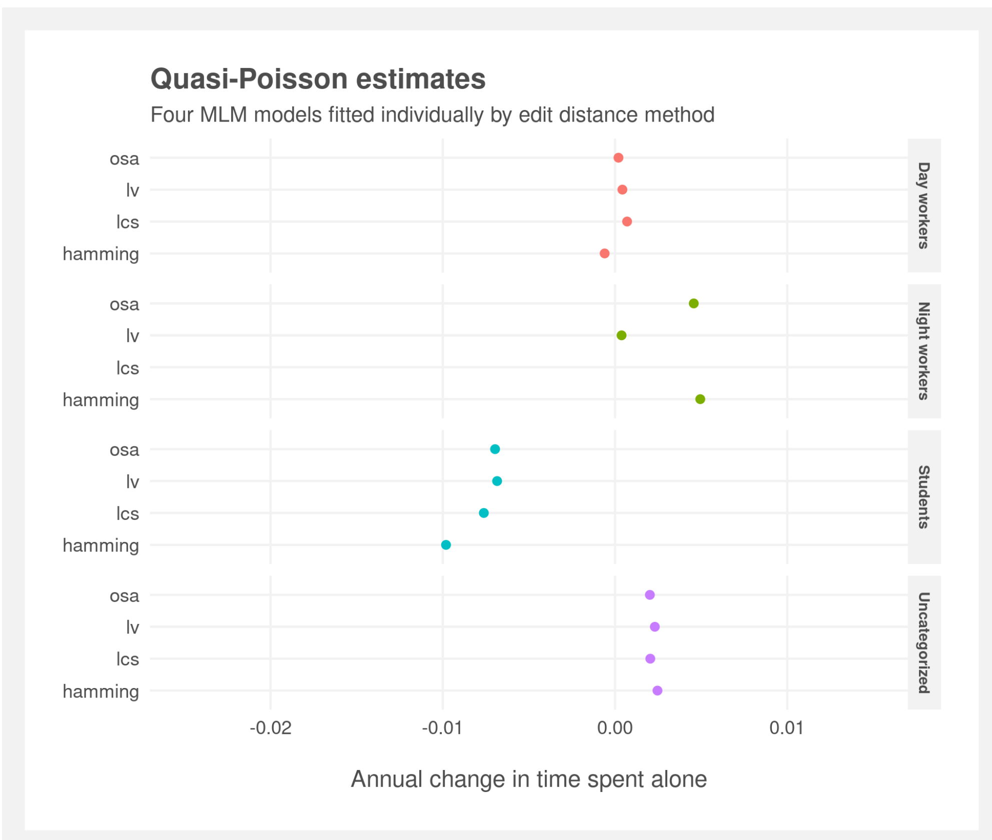

Figure 13: All edit distance measures: Multilevel quasi-Poisson estimates

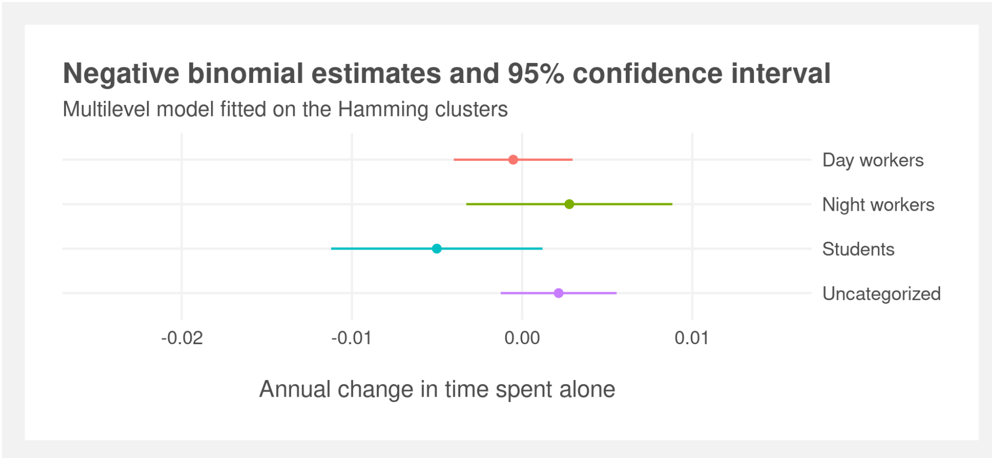

Figure 14: Hamming edit distance: Multilevel negative binomial effects

Associated demographics

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Mattis enim ut tellus elementum sagittis vitae et. Odio pellentesque diam volutpat commodo sed egestas egestas fringilla. Consectetur lorem donec massa sapien faucibus. Euismod nisi porta lorem mollis aliquam ut porttitor leo a. Metus dictum at tempor commodo. Aliquam etiam erat velit scelerisque in dictum non consectetur. Volutpat odio facilisis mauris sit amet massa vitae tortor condimentum. Nibh venenatis cras sed felis eget velit.

The associated demographics of these clusters skew ...

Figure 15: Densities of age by cluster and method

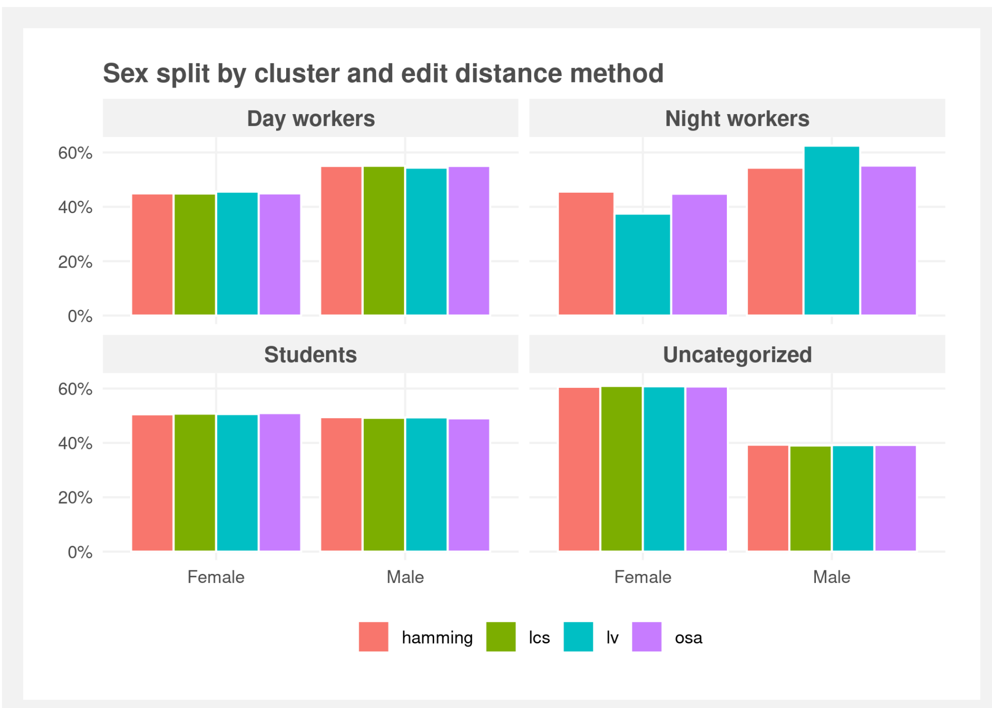

Figure 16: Sex split by cluster and method

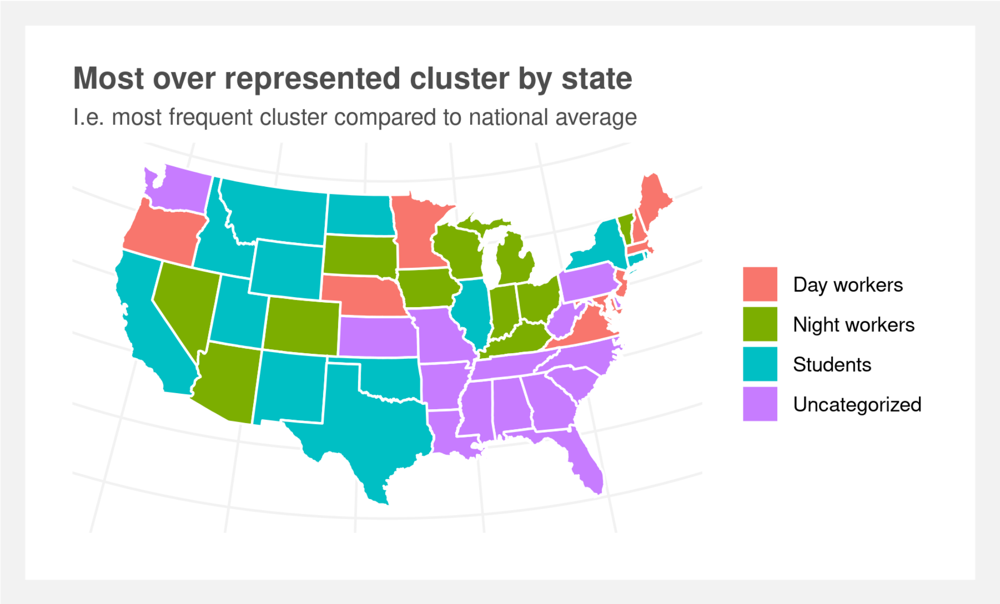

Figure 17: Most over represented cluster by state

explore other cluster solutions here that identify the elderly? The Hamming, Levenshtein, and OSA edit distance measures indicate that five or possibly six clusters are reasonable (Figure 6). The six-cluster solution . . . . contains clusters can be categorized as . . .

Conclusion

Facilisi cras fermentum odio eu. Quis enim lobortis scelerisque fermentum dui faucibus in. Commodo elit at imperdiet dui. Mi quis hendrerit dolor magna eget. Scelerisque purus semper eget duis at tellus at. Aliquet bibendum enim facilisis gravida neque. Sagittis eu volutpat odio facilisis mauris sit amet massa. Aliquet lectus proin nibh nisl condimentum. A cras semper auctor neque vitae. Lobortis scelerisque fermentum dui faucibus in ornare quam viverra orci. Est lorem ipsum dolor sit amet consectetur adipiscing elit.

Eleifend mi in nulla posuere sollicitudin aliquam ultrices sagittis. Ultricies mi quis hendrerit dolor magna eget est. Pharetra diam sit amet nisl suscipit adipiscing. Donec pretium vulputate sapien nec. Arcu odio ut sem nulla pharetra. Dolor sit amet consectetur adipiscing elit pellentesque habitant morbi.

Appendix

Table 2: Activity aggregation mapping

Activity code	Description
t0101.*	Sleep
t010[2-9].*	Personal Care
t019.*	Personal Care
t1801.*	Personal Care
t02.*	Household Activities
t1802.*	Household Activities
t03.*	Caring For Household Member
t1803.*	Caring For Household Member
t04.*	Caring For Nonhousehold Members
t1804.*	Caring For Nonhousehold Members
t05.*	Work
t1805.*	Work
t06.*	Education
t1806.*	Education
t07.*	Consumer Purchases
t1807.*	Consumer Purchases
t08.*	Professional & Personal Care Services
t1808.*	Professional & Personal Care Services
t09.*	Other
t1809.*	Other
t10.*	Other
t1810.*	Other
t11.*	Eating and Drinking
t1811.*	Eating and Drinking
t12.*	Socializing, Relaxing, and Leisure
t1812.*	Socializing, Relaxing, and Leisure
t13.*	Sports, Exercise, and Recreation
t1813.*	Sports, Exercise, and Recreation
t14.*	Religious and Spiritual
t1814.*	Religious and Spiritual
t15.*	Volunteer
t1815.*	Volunteer
t16.*	Other
t1816.*	Other
t1818.*	Other
t1819.*	Other
t189.*	Other
t50.*	Other

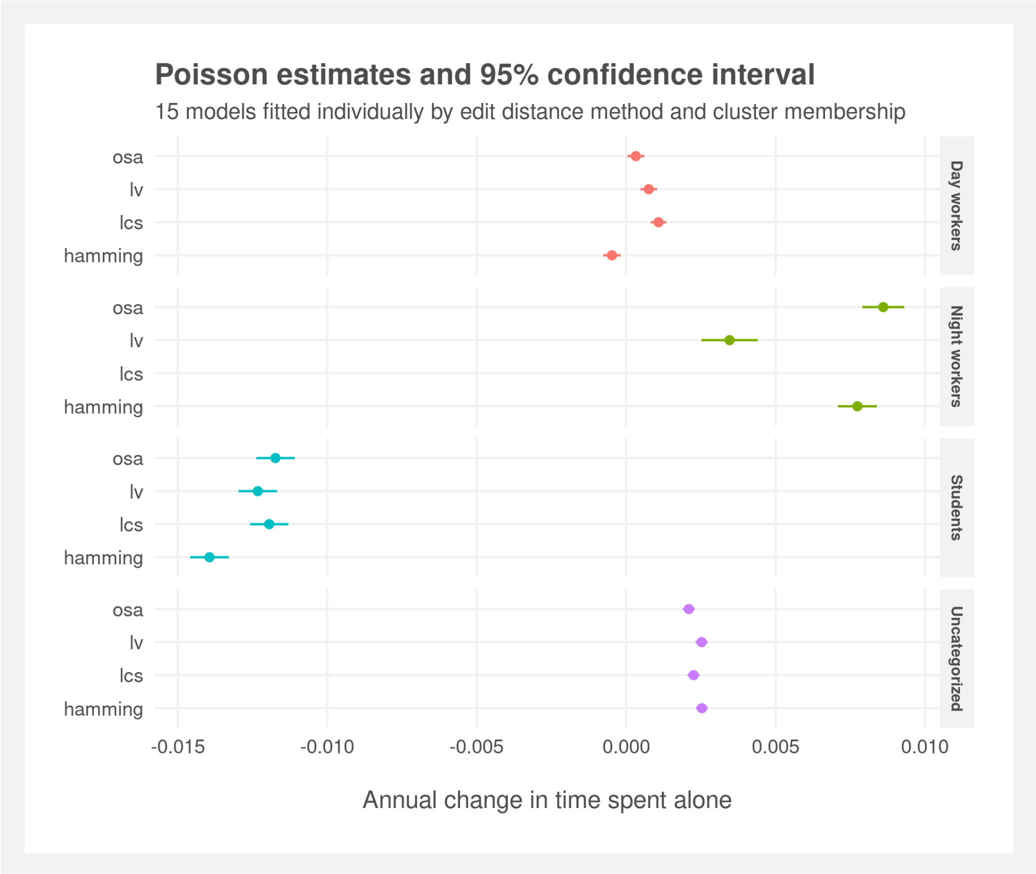

Figure 18: Poisson single-level effects

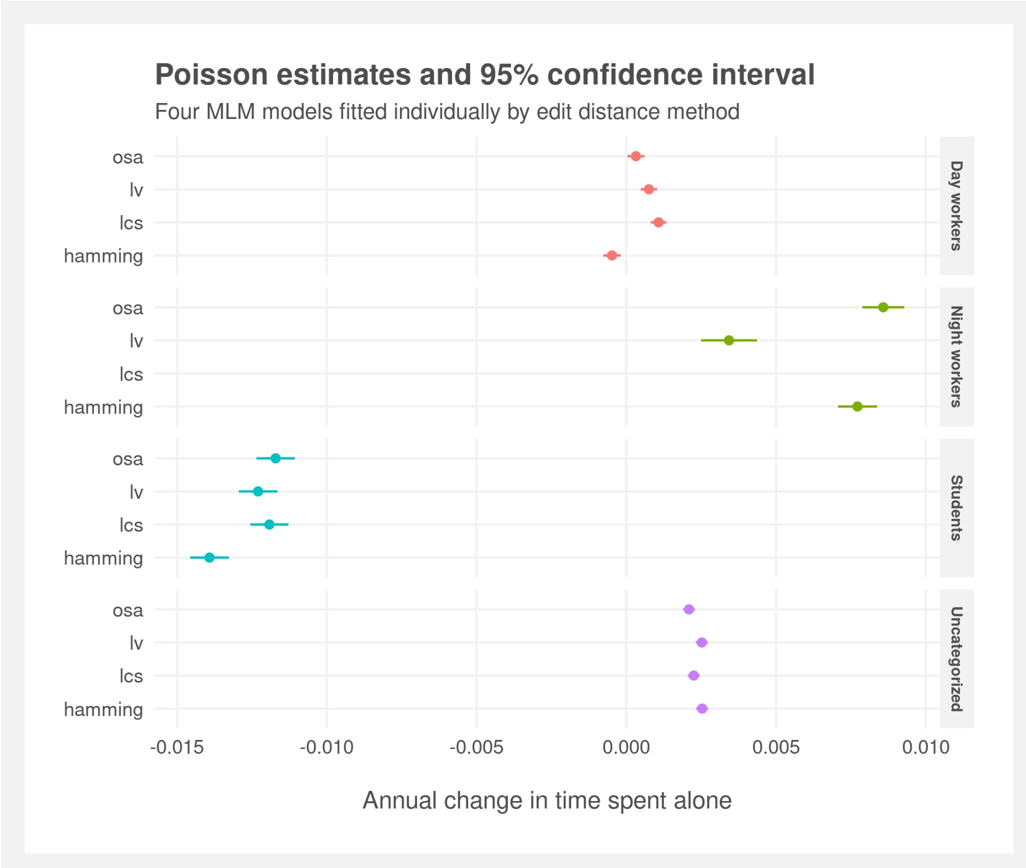

Figure 19: Poisson MLM effects

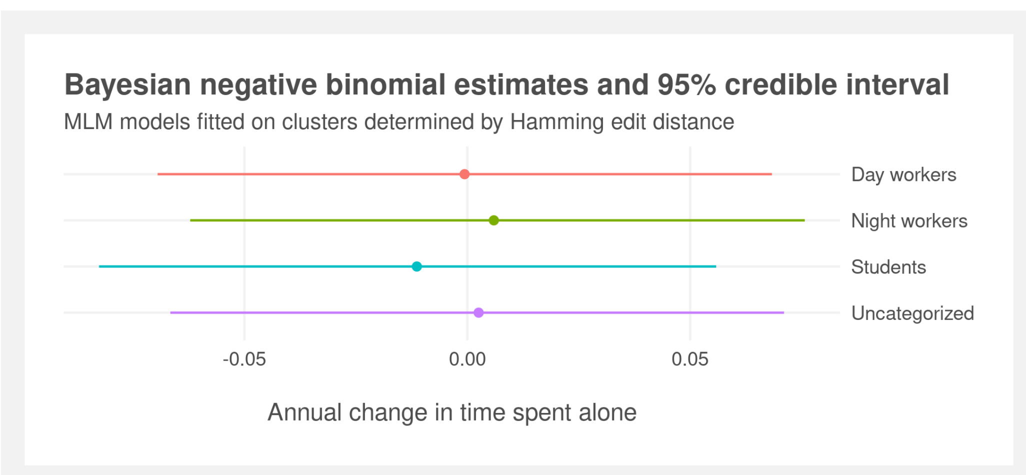

Figure 20: Bayesian negative binomial MLM effects