Overview

This document explains how to replicate the empirical findings reported in the figures and tables of the paper, titled “Living in Sync with the Sun: Sleep and Mental Health Implications of Circadian Misalignment”, and the empirical findings reported in the paper’s online appendix. The replication work uses the following file structure:

```
project_dst_replication
├── code
│   ├── management
│   ├── analysis
│   │   ├── figures
│   │   │   ├── appendix
│   │   ├── tables
│   │   │   ├── appendix
│   ├── user_written_packages
├── data_raw
│   ├── acs
│   ├── counties
│   ├── places
│   ├── solar
│   ├── time_zones
└── data_clean
    ├── acs
    ├── counties
    ├── places
    ├── solar
```

Data Availability

Summary

- [x] **All data are publicly available.**
- [ ] **Some data cannot be made publicly available.**
- [ ] **No data can be made publicly available.**

Data Sources

- PLACES from the CDC (https://www.cdc.gov/places/):
  - Data portal (https://chronicdata.cdc.gov/browse?category=500+Cities+%26+Places)
  - 2020 release (https://chronicdata.cdc.gov/browse?q=2020%20release) includes 27 outcomes based on the 2017 and 2018 BRFSS data:
    - “PLACES: Local Data for Better Health 2018” layer (https://lu.maps.arcgis.com/home/item.html?id=8eca985039464f4d83467b8f6aeb1320) for the 2020 release was imported into ArcGIS Pro (version 2.9.3) from Esri’s Living Atlas by Muzhe Yang on 06/06/2022.
    - Census tract level dataset named “places_2020_release_census_tracts” was exported to the ArcGIS Pro’s file geodatabase from the ArcGIS Pro’s sublayer named “Tracts” of the “PLACES: Local Data for Better Health 2018” layer on 06/06/2022 by Muzhe Yang. This dataset was then exported from the ArcGIS Pro as an EXCEL dataset named “places_2020_release_census_tracts.xlsx” saved under this folder named `data_raw\places\` .
    - Census tract level feature layer in ArcGIS Pro named “places_2020_release_census_tracts_layer_projected” was exported to the ArcGIS Pro’s file geodatabase from the sublayer named “Tracts” of the “PLACES: Local

Data for Better Health 2018" layer on 06/08/2022 by Muzhe Yang. This layer uses "WGS_1984_Web_Mercator_Auxiliary_Sphere" as the output coordinate system.

    - 2021 release (https://chronicdata.cdc.gov/browse?q=2021%20release), which is the most recent release as of 12/2021, includes 29 outcomes based on the 2018 and 2019 BRFSS data:
        - "PLACES: Local Data for Better Health" layer (https://lu.maps.arcgis.com/home/item.html?id=3b7221d4e47740cab9235b839fa55cd7) for the 2021 release was imported into ArcGIS Pro (version 2.9.3) from Esri's Living Atlas by Muzhe Yang on 06/15/2022.
        - Census tract level dataset named "places_2021_release_census_tracts" was exported to the ArcGIS Pro's file geodatabase from the ArcGIS Pro's sublayer named "Tracts" of the "PLACES: Local Data for Better Health" layer on 06/15/2022 by Muzhe Yang. This dataset was then exported from the ArcGIS Pro as an EXCEL dataset named "places_2021_release_census_tracts.xlsx" saved under this folder named `data_raw\places\`.
        - Census tract level feature layer in ArcGIS Pro named "places_2021_release_census_tracts_layer_projected" was exported to the ArcGIS Pro's file geodatabase from the sublayer named "Tracts" of the "PLACES: Local Data for Better Health" layer on 06/15/2022 by Muzhe Yang. This layer uses "WGS_1984_Web_Mercator_Auxiliary_Sphere" as the output coordinate system.
- Time zones:
    - The U.S. time zone shapefiles (https://data-usdot.opendata.arcgis.com/datasets/usdot::time-zones/about) was obtained from the USDOT BTS by Muzhe Yang:
        - Clicked "Download" then went to "Shapefile" then clicked "Download Options" then selected "Generate new download with latest data" or "Download file previously generated on June 5, 2022, 08:34" (the first download was done by Muzhe Yang on 06/05/2022, 8:34 AM).
        - After the download, the "Time_Zones.shp" (together with the .cpg, .dbf, .prj and .shx) saved under `data_raw\time_zones\Time_Zones\` was imported into the ArcGIS Pro's file geodatabase and renamed to "us_time_zones" on 06/05/2022 by Muzhe Yang.
- World latitudes and longitudes:
    - The "World Latitude and Longitude Grids" layer (https://lu.maps.arcgis.com/home/item.html?id=ece08608f53949a4a4ee827fd5c30da1) in ArcGIS Pro was obtained from Esri's Living Atlas on 06/05/2022 by Muzhe Yang.
    - The "World GeoReference Lines" layer (https://lu.maps.arcgis.com/home/item.html?id=68f844f9efd04df6a4636e0f410ac072) in ArcGIS Pro was obtained from Esri's Living Atlas on 06/05/2022 by Muzhe Yang.
- Census tracts:
    - The census tracts boundary layer named "us_census_tracts" in ArcGIS Pro was obtained from here (https://esri.maps.arcgis.com/home/item.html?id=20f5d275113e4066bf311236d9dcc3d4) on 06/11/2022 by Muzhe Yang:
        - "This layer presents the USA 2020 Census tract boundaries of the United States in the 50 states and the District of Columbia."
    - Census tract level demographic variables were obtained from the American Community Survey (ACS) through Esri's Living Atlas by Muzhe Yang:
        - As of 06/11/2022, only 2010–2014 ACS and 2016–2020 ACS data were available.
        - We used 2016–2020 ACS data for this project. This is reasonable because many heath outcomes in the PLACES data used for this project were measured in 2018, and 2018 is right in the middle of 2016–2020.
            - The ACS 5-year estimates should be used for getting information about each census tract. See "When to Use 1-year or 5-year Estimates" (https://www.census.gov/programs-surveys/acs/guidance/estimates.html).
            - ACS's "Table IDs Explained" (https://www.census.gov/programs-surveys/acs/data/data-tables/table-ids-explained.html)
            - About the "margin of error": "90% confidence level is the Census Bureau Standard" (source (https://www.census.gov/content/dam/Census/programs-surveys/acs/guidance/training-presentations/20180418_MOE.pdf))
        - The following 6 feature layers were imported into ArcGIS Pro on 06/11/2022 by Muzhe Yang:
            - "ACS Population and Housing Basics - Boundaries" (https://lu.maps.arcgis.com/home/item.html?id=9ae2ecb9ad4c4a0fbdafa5474f7cbb5e)

        - “ACS Educational Attainment Variables - Boundaries” (https://lu.maps.arcgis.com/home/item.html?id=84e3022a376e41feb4dd8addf25835a3)
        - “ACS Marital Status Variables - Boundaries” (https://lu.maps.arcgis.com/home/item.html?id=165fb3ab31f444a69645029b04579e30)
        - “ACS Household Size Variables - Boundaries” (https://lu.maps.arcgis.com/home/item.html?id=388cebd5976e49faa77af91a5d73dfee)
        - “ACS Health Insurance Coverage Variables - Boundaries” (https://lu.maps.arcgis.com/home/item.html?id=a1574f4bb84f4da78b60fa0c8616eaa1)
        - “ACS Employment Status Variables - Boundaries” (https://lu.maps.arcgis.com/home/item.html?id=a93ad1fcf4a64580b7dbfa3758252ef0)
    - The attribute tables containing the census tract level data associated with the 6 feature layers described above were exported to the ArcGIS Pro’s file geodatabase on 06/11/2022 by Muzhe Yang. Those data were then exported from the ArcGIS Pro as EXCEL files in the .xlsx format (note: the .xls format cannot be used because of the .xls format’s limitation on the input file’s row numbers) saved under this folder named `data_raw\acs\`.
- Counties:
    - Latitude and longitude of the centroid of each U.S. county: Browsed this website (https://www.census.gov/), then browsed by Topics, then Geography, then Reference Files, then Gazetter Files (https://www.census.gov/geographies/reference-files/time-series/geo/gazetteer-files.html). Muzhe Yang selected 2020 to download “Gaz_counties_national.zip” on 06/19/2022, then unzipped it, and then renamed it to “Gaz_counties_national_2020.txt” and saved it under this folder `data_raw\counties\`.
        - The year 2020 was selected for two reasons: (1) The year selected must be later than 2015 because of a county’s name and the associated FIPS code change that happened in 2015: “Oglala Lakota County (known as Shannon County until May 2015) is a county in southwestern South Dakota, United States.” (information source (https://en.wikipedia.org/wiki/Oglala_Lakota_County,_South_Dakota)); (2) The CDC’s PLACES started in 2020.
        - For descriptions of variables, see here (https://www.census.gov/programs-surveys/geography/technical-documentation/records-layout/gaz-record-layouts.2020.html).
        - For additional detailed description of variables, see here (https://www.census.gov/programs-surveys/geography/technical-documentation/complete-technical-documentation/tiger-geodatabase-file.html) and click “2020 TIGER Geodatabase Documentation.pdf”.
        - The latitude and longitude of the so-called “internal point” are named “INTPTLAT” and “INTPTLONG” in the original data provided by the Census Bureau. The “internal point” is the geographic center.
            - For population center, see here (https://www.census.gov/geographies/reference-files/time-series/geo/centers-population.html). In this case and in the original data provided by the Census Bureau, the latitude of the population center is called “LATITUDE”, and the longitude of the population center is called “LONGITUDE”.

Data Processing

Note: The “file geodatabase” mentioned in the steps below is the one used by the ArcGIS Pro software. The version of the ArcGIS Pro used for data processing conducted in June 2022 is 2.9.3. Muzhe Yang did the following data processing work. The steps are listed in logical order, not necessarily in the order of dates when the data work was done:

- Step 1. Coordinate systems:
    - Projection on the fly uses the coordinate systems (GCS and PCS) set by the the “PLACES: Local Data for Better Health 2018” layer: “WGS_1984_Web_Mercator” (PCS) and “GCS_WGS_1984_Major_Auxiliary_Sphere” (GCS).
    - The “WGS_1984_Web_Mercator” doesn’t seem to be a standard name, which caused ArcGIS Pro’s “project” tool unable to work. To solve this problem, Muzhe picked “WGS 1984 Web Mercator (auxiliary sphere)” to be the coordinate system, which should be applied to all layers that are used for spatial analyses. The names of these layers have “_projected” at the end.
- Step 2. Created three boundaries (i.e., line feature layers) by tracing the “us_time_zones” layer’s polygons (while relying on ArcGIS Pro’s projection on the fly): “boundary_eastern_central_projected”, “boundary_central_mountain_projected”, and “boundary_mountain_pacific_projected”.

- Step 3. Created three line feature layers for the ideal time zones boundaries by using the “World Latitude and Longitude Grids” for these three longitudes, respectively: -82.5 degrees, -97.5 degrees, and -112.5 degrees.
- Step 4. Created two line feature layers using the “World Latitude and Longitude Grids” for these two latitudes, respectively: 34 degrees and 40 degrees.
- Step 5. Created the point feature class of the centroid of each census tract included in the 2020 release of the PLACES. The associated feature layer is named “places_2020_release_census_tracts_centroid_projected”. There are two steps in the creation process.
    - Step (i), which is “Table to Point Feature Class”, uses “GCS_WGS_1984” as the coordinate system and “places_2020_release_census_tracts” in the file geodatabase as the input table.
    - Step (ii), which is “Project”, uses “WGS_1984_Web_Mercator_Auxiliary_Sphere” as the output coordinate system.
    - Note: The PLACES 2020 release and the PLACES 2021 release include the same census tracts.
- Step 6. Spatial-intersected the “places_2020_release_census_tracts_centroid_projected” layer with the “us_time_zone” layer, so that each census tract has the time zone variable. This was done on 06/10/2022 by using the “Pairwise Intersect” tool since only two feature layers are used for the intersection. The result is the layer and the point feature class (with irrelevant fields joined from the “us_time_zones” layer deleted) named “places_2020_release_ct_centroid_proj_intersect”.
    - There are 72,337 census tracts in the PLACES 2020 release data. After the spatial intersection, there are 72,332 census tracts. The census tracts that cannot be found in any of the time zone polygons in the “us_time_zones” layer were dropped.
- Step 7. Exported the “places_2020_release_ct_centroid_proj_intersect” layer to the following three feature layers on 06/10/2022: “places_2020_release_eastern_central”, “places_2020_release_central_mountain”, and “places_2020_release_mountain_pacific”.
- Step 8. Calculated the distance between each census tract centroid and the boundary of two adjacent time zones (on 06/10/2022), using the “Near” tool and selecting the geodesic method, for the three feature layers explained above, respectively.
- Step 9. The attribute tables of these three feature layers—“places_2020_release_eastern_central”, “places_2020_release_central_mountain”, and “places_2020_release_mountain_pacific”—were exported as EXCEL files in the .xlsx format on 06/12/2022 (note: the .xls format cannot be used because of the .xls format’s limitation on the input file’s row numbers) saved under this folder named `data_raw\places\` : “places_2020_release_eastern_central.xlsx”, “places_2020_release_central_mountain.xlsx”, and “places_2020_release_mountain_pacific.xlsx”.
- Step 10. Used the “Join Field” geoprocessing tool in ArcGIS Pro to add the “Zone” field in the “places_2000_release_ct_centroid_proj_intersect” layer’s attribute table to the “places_2021_release_census_tracts_layer_projected” layer’s attribute table (i.e., to add the time zone information) on 06/15/2022.
    - There are 72,337 observations in the “places_2021_release_census_tracts_layer_projected” layer’s attribute table. Five of these observations have no information on time zones. This is correct, and it is the result of the spatial interaction explained above. In the end, there are 72,332 census tracts that have time zone information.
- Step 11. Exported the “places_2021_release_projected” layer to the following three feature layers on 06/15/2022: “places_2021_release_eastern_central”, “places_2021_release_central_mountain”, and “places_2021_release_mountain_pacific”.
- Step 12. Used the “Join Field” geoprocessing tool in ArcGIS Pro to add the “NEAR_DIST” field in the attribute table of the “places_2020_release_X” layer to the attribute table of the “places_2021_release_X” layer (i.e., to add the information on the distance between each census tract centroid and the boundary of two adjacent time zones), where X = “eastern_central”, “central_mountain”, and “mountain_pacific” on 06/15/2022.
- Step 13. The attribute tables of these three feature layers—“places_2021_release_eastern_central”, “places_2021_release_central_mountain”, and “places_2021_release_mountain_pacific”—were exported as EXCEL files in the .xlsx format on 06/15/2022 (note: the .xls format cannot be used because of the .xls format’s limitation on the input file’s row numbers) saved under this folder named `data_raw\places\` : “places_2021_release_eastern_central.xlsx”, “places_2021_release_central_mountain.xlsx”, and “places_2021_release_mountain_pacific.xlsx”.
- Step 14. The attribute tables of the census tract level data associated with the 6 feature layers described in the “Census tracts” section were exported to the file geodatabase on 06/11/2022.
    - Exporting these data to the file geodatabase is necessary for two reasons: one is preserving the data as of 06/11/2022; and the other is making those data become user-owned so that data wrangling is allowed.

    - Those data were then exported as EXCEL files in the .xlsx format saved under this folder named `data_raw\acs\`: "acs_pop_and_housing_basics.xlsx", "acs_educ_attainment.xlsx", "acs_marital_status.xlsx", "acs_hh_size.xlsx", "acs_health_ins_coverage.xlsx", and "acs_employ_status.xlsx".
- Step 15. Calculated the sunrise time, sunset time, and the sunlight duration (i.e., sunset time minus sunrise time) for each county's centroid among the lower 48 states (plus DC), and for the year 2018 since most health outcomes in the PLACES data were measured in that year. The Stata code "solar_calculator.ado" written by Jeffrey Shrader was downloaded from here (https://github.com/jshrader/astronomical_algorithms).
    - This code is the "Stata version of NOAA's implementation of Meeus' sunset time calculation." This code was used in this RESTAT publication (https://doi.org/10.1162/rest_a_00746).
    - The formulas used in this Stata code follow the EXCEL files such as "NOAA_Solar_Calculations_year.xls" posted at this website (https://gml.noaa.gov/grad/solcalc/calcdetails.html). All the formulas can be seen in the cells of those EXCEL files.
        - The EXCEL file including all dates in the year 2018 was obtained from "NOAA_Solar_Calculations_year.xls" by setting the "Year" (in that EXCEL file) to 2018.
    - The sunlight duration (i.e., sunset time minus sunrise time) in the EXCEL file is measured in minutes. But, the Stata code has the "Hours of sunlight" label for the "sunlight_duration" variable. This is a mistake in the Stata code, and the author of the Stata code made a correction on this website (https://github.com/jshrader/astronomical_algorithms): "sunlight_duration - In minutes (despite what the variable label says)".

Dataset List

- All original data are under the `data_raw` folder.
- All cleaned data are under the `data_clean` folder.
- All script files using the original data to generate the cleaned data, and the associated log files, are under the `code\management` folder.

Computational Requirements

Computer and the Operating System (OS)

The code was last run on a desktop Dell computer with the following specifications:

- Processor: 13th Gen Intel(R) Core(TM) i9-13900 2.00 GHz
- RAM: 64.0 GB
- System type: 64-bit operating system, x64-based processor
- OS: Microsoft Windows
    - Edition: 11 Pro
    - Version: 23H2
    - OS build: 22631.4751

Softwares

Two softwares used by this paper are:

- Stata/MP 18 8-score: all user-written packages required for the replication work are under the `code\user_written_packages\` folder.
- ArcGIS Pro, version 2.9.3.

Controlled Randomness

Random-number seeds were used for replicability. See script files for details.

Runtime

Each log file recorded the start time and the end time of a program's execution.

Descriptions of Code/Program/Script Files

All script files, their associated log files and output files are under the `code` folder.

- `dst-profile.do` : Must run this do file first, in order to set the working directory and to add/remove a directory to/from Stata's search path (i.e., to set up a virtual environment in Stata).
- `dst-main.do` : This do file generates the final dataset used for creating the figures and tables of this paper and the paper's online appendix. All do files used by `dst-main.do` are under the `code\management\` folder.
- Run each do file under the `code\analysis\` folder to replicate each figure and each table in the paper and in the paper's online appendix.

The provided script files replicate all figures and all tables in the paper and in the paper's online appendix, except the following:

- In the paper:
  - Figure 1 and Figure 2: They were created by ArcGIS Pro.
  - Table 1: It was created by EXCEL.
- In the paper's online appendix:
  - Appendix Figure A1: It was created by ArcGIS Pro.
  - Appendix Table A1: It was created by EXCEL.