

# Data Provenance - John Woodward

---

## Data

### U.S. states by purchase power

Description: This Wikipedia page provides a ranked list of U.S. states by their adjusted per capita personal income, which considers the cost of living differences between states.

Date of Download: 2023-04-21

Source:

[https://en.wikipedia.org/wiki/List\\_of\\_U.S.\\_states\\_by\\_adjusted\\_per\\_capita\\_personal\\_income](https://en.wikipedia.org/wiki/List_of_U.S._states_by_adjusted_per_capita_personal_income)

Terms of Service:

[https://foundation.wikimedia.org/wiki/Policy:Terms\\_of\\_Use/en#4.\\_Refraining\\_from\\_Certain\\_Activities](https://foundation.wikimedia.org/wiki/Policy:Terms_of_Use/en#4._Refraining_from_Certain_Activities)

License: Creative Commons Attribution-ShareAlike 3.0 Unported License (CC BY-SA 3.0)

### U.S. states and abbreviations

Description: This Wikipedia page lists all 50 states, with brief descriptions and related information.

Date of Download: 2023-04-21

Source: [https://en.wikipedia.org/wiki/List\\_of\\_states\\_and\\_territories\\_of\\_the\\_United\\_States](https://en.wikipedia.org/wiki/List_of_states_and_territories_of_the_United_States)

Terms of Service:

[https://foundation.wikimedia.org/wiki/Policy:Terms\\_of\\_Use/en#4.\\_Refraining\\_from\\_Certain\\_Activities](https://foundation.wikimedia.org/wiki/Policy:Terms_of_Use/en#4._Refraining_from_Certain_Activities)

License: Creative Commons Attribution-ShareAlike 3.0 Unported License (CC BY-SA 3.0)

### U.S. states shp files

Description: This government URL provided the shp files for my needed state visuals.

Date of Download: 2023-04-30

Source: [https://www2.census.gov/geo/tiger/TIGER2020/STATE/tl\\_2020\\_us\\_state.zip](https://www2.census.gov/geo/tiger/TIGER2020/STATE/tl_2020_us_state.zip)

License: Creative Commons Attribution-ShareAlike 3.0 Unported License (CC BY-SA 3.0)

## U.S. city latitude and longitude

Description: The GeoNames geographical database covers all countries and contains over eleven million placenames available for download free of charge. The US.zip file houses a TXT file containing millions of U.S. cities and information, including the latitude and longitude.

Date of Download: 2023-04-29

Source: <http://download.geonames.org/export/dump/>

License: Creative Commons Attribution-ShareAlike 4.0 Unported License (CC BY-SA 4.0)

## Job Search results

Description: I used the URL provided to conduct job searches on LinkedIn, with the search criteria being: (cplex OR "linear programming" OR gurobi) and (python sql).

Following this, a Lix browser plug-in was downloaded

<https://chrome.google.com/webstore/detail/lix-linkedin-scraping-ema/ceplokfhfeekddamgoaojabdhkkgafnk>, registered at <https://lix-it.com/login>, and linked with an active LinkedIn account.

Subsequently, I applied the Lix-it tool to obtain high-level meta-data for the job listings (e.g., Organization, Job Title).

Lastly, I manually downloaded the lower-level job description data by hiring a Fiverr contractor for \$50 to comply with the terms and conditions prohibiting automated web scraping. The process took four hours of manual work, was one-time, and performed non-commercially.

In retrospect, the Lix tool gathering metadata may have skirted terms of service for LinkedIn, but according to the additional context provided, this is considered acceptable.

Date of Download: 2023-04-11

Source:

[https://www.linkedin.com/jobs/search/?currentJobId=3518862183&distance=25&geoId=103644278&keywords=\(cplex%20OR%20%E2%80%9Clinear%20programming%E2%80%9D%20OR%20gurobi\)%20and%20\(python%20sql\)](https://www.linkedin.com/jobs/search/?currentJobId=3518862183&distance=25&geoId=103644278&keywords=(cplex%20OR%20%E2%80%9Clinear%20programming%E2%80%9D%20OR%20gurobi)%20and%20(python%20sql))

Terms of Service: <https://www.linkedin.com/legal/user-agreement#dos>

License: N/A

Additional Context: <https://www.zdnet.com/article/court-rules-that-data-scraping-is-legal-in-linkedin-appeal/>  
<https://www.eff.org/deeplinks/2020/04/federal-judge-rules-it-not-crime-violate->

websites-terms-service

<https://www.eff.org/deeplinks/2022/04/scraping-public-websites-still-isnt-crime-court-appeals-declares>