**Group 8** **members**

Dalton Driver, Andrew Kliever, Drew Johnson, Martin Rybicki, and Sarah Nystrom.

**HYPOTHESES**

**Hypothesis (H1)**

The geographical location of US Army bases is positively correlated with poverty rates in surrounding areas, such that areas with US Army bases exhibit higher poverty rates compared to areas without.

**Null Hypothesis (H0)**

There is no significant correlation between the geographical location of US Army bases and the poverty rates in surrounding areas.

**Alternative Hypothesis (H2)**

There is a significant positive correlation between the geographical location of US Army bases and the poverty rates in surrounding areas, indicating that areas with US Army bases have higher poverty rates than those without.

**PROJECT INFORMTION**

**Project Questions**

1. Does base size impact the poverty rate?
2. Does the geographical location of the forts impact the poverty rate?
3. Does the state the fort is in impact the poverty rate?
4. Does the local population size impact the poverty rate?

**Limiting Parameters**

* We limited the project to the US Army branch only.
* We limited the bases to US Army forts (larges bases) only.
* We located the nearest town or city for poverty rates as some bases are so large they cover one or multiple zip codes.

**Data Sources**

We pulled information for data such as population, poverty rates, and household income from the US Census Bureau (data.densus.gov). We pulled information from the US Department of Defense (defense.gov) and from Military OneSource (militaryonesource.gov) for data such as base information and locations. We also used Google and ChatGPT to validate some of the data pulled, to answer questions that arose, and to read articles related to our topic and findings.

**Data Cleansing**

* The csv file from the US Census Bureau had random symbols, like exclamation marks, that we had to do a find/replace to remove.
* We deleted some columns, such as Median Age, that we deemed unnecessary for our purposes.
* When we ran the code, we got a strange output for 5 of the 27 base zip codes. It was a string of 6s. Through Google and ChatGPT, we learned these bases were so large, they covered an entire zip code or multiple zip codes.
* The US Census Bureau listed the 2023 poverty rate as 11.4 percent. We looked through the past years and saw the rate fluctuated between almost 13 percent and 11 percent. To provide the best viewpoint into the data, we first ran the data visualization using the 13 percent, then visualized the data using the 2023 rate of 11.4 percent.

**Visualizations**

1. *Geoplot:* Base locations and sizes
2. *Pie chart:* Forts by Poverty Rate
3. *Vertical bar graph:* Poverty Rate by Fort
4. *Horizontal bar graph*: Poverty Rate by State
5. *Bubble chart:* Population vs Poverty Rate with Household Income
6. *Scatter with linear:* Population Size vs Poverty Rate around Bases
7. *Scatter with linear:* Population vs Household Income around Bases
8. *Heatmap*: Poverty Rate by Binned Population and Household Income
9. *Geoplot:* Geographic Distribution of Poverty Rates around Bases

**ANALYSIS**

Our data with above-stated limitations produced 27 US Army bases.

Using geographical delineations, the dispersion of bases is as follows (states with bases are **bolded**):

|  |  |  |
| --- | --- | --- |
| **Geographic location** | **States** | **Bases** |
| West | **WA.** OR. **CA.** MT. ID. NV. UT. WY. **CO.** AK. HI. | In 3 of 11 states |
| Southwest | **AZ.** NM. **TX. OK.** | In 3 of 4 states |
| Midwest | ND. SD. NE. **KS.** MN. IA. **MO. WI.** IL. IN. MI. OH. | In 3 of 12 states |
| Southeast | AR. **LA.** MS. **AL.** TN. **KY.** WV. **VA. NC. SC. GA.** FL. | In 7 of 12 states |
| Northeast | DE. **MD.** NJ. PA. RI. CT. MA. NH. **NY.** VT. ME. | In 2 of 11 states |

***Map of Base Locations and Sizes***

When viewing the visualization of the **map of base locations and sizes**, the larger bases are predominantly in the more rural or less-populated areas of the country. This could indicate a reason for a correlation between base location (whether urban or rural) and poverty rates.

***Poverty Rate by Fort***

When viewing the two visualizations of **poverty rate by fort**, only 8 of the 27 forts are below the national poverty line:

|  |  |  |
| --- | --- | --- |
| **Base name** | **Base location** | **Poverty ranking on chart** |
| Fort Leavenworth | Kansas | 20 of 27 |
| Fort Hamilton | New York | 21 of 27 |
| Fort Knox | Kentucky | 22 of 27 |
| Fort McCoy | Wisconsin | 23 of 27 |
| Fort Detrick | Maryland | 24 of 27 |
| Fort Meyer | Virginia | 25 of 27 |
| Fort George G. Meade | Maryland | 26 of 27 |
| Fort Belvoir | Virginia | 27 of 27 |

The proximity of the four bases with the lowest poverty rates to Washington D.C. should be considered when further investigating the correlation between base location and poverty rate. The demographic diversity, both historical and current, of these areas should be considered as well.

Whereas Fort Belvoir and Fort George G. Meade are outliers on the low poverty rate, Fort Jackson and Fort Bliss appear on the opposite side of the bar graph indicating abnormally high poverty rates:

|  |  |  |
| --- | --- | --- |
| **Base name** | **Base location** | **Poverty ranking on chart** |
| Fort Bliss | Texas | 1 of 27 |
| Fort Jackson | South Carolina | 2 of 27 |

Both Texas and South Carolina have urban and rural areas. Looking into the locations around the base could correlate to the high poverty rates.

Also, it is common knowledge Texas has a diverse population with a high population of Hispanic and Latino makeup and South Carolina has a high percentage of Black and African American populations. Future data could analyze the correlation between historically marginalized and/or economically disadvantages populations in relation to base locations.

***Poverty Rate by State***

The visualization of **poverty rate by state** for fort locations shows South Carolina as the top state with highest poverty rate by fort location. Texas is the second. This is notable because South Carolina only has one base in this study (Fort Jackson) Texas has two bases (Fort Bliss and Fort Cavazos). This is notable because Fort Bliss (Texas) is the base at the top of the poverty ranking by base. Future data could investigate the specific poverty rates of these three bases.

***Population vs. Poverty Rate***

In this bubble chart, population is regarding the town closest to the fort. Another addition is the national poverty rate as the red dashed line added to provide context. In addition to the axes, the bubbles represent a third dimension of data; the household income found in the towns. The larger the bubbles, the higher the per family income in the cities.

With this chart we see the beginnings of a possible correlation between population, poverty rate, and household income. It does seem that as population increases, the poverty rate of the city nearest to the fort decreases slightly, with the 15k population segment and 40k segment showing the lowest poverty rates and highest household incomes. This seems to not be the case when the population of the cities go beyond 50k.

***Scatter plot 1***

We must break down the first chart in sub-parts and apply some measurement to these observations. Here population size and poverty rate are being compared with the national poverty rate thrown in for context with linear regression showing a -0.39 negative correlation, meaning, as population increases the poverty rate decreases. The correlation is considered weak due to falling between 0.2 and 0.39. Please note the outliers, particularly at the top left which is the city of El Paso which leans heavily into the "smaller population = poverty rate" observation.

***Scatter plot 2***

We then look the next chart which is population size and household income, and we find a positive correlation between the two of 0.20, which makes it a weak but still existing correlation. As the population of the cities increase the household income increases. Please note the outliers, particularly the top left datapoint of Fort Belvoir with its city being Alexandria. There are a few more outliers as the population increases, particularly at the 40k mark, but not as extreme. If Alexandria is taken out, the positive correlation would be higher.

***Heatmap with Bins***

The heatmap shows us population and household income and poverty rate from a different perspective, with household income now holding the x-axis, population the y, and the poverty rate being represented by the color with yellow being higher poverty rate.

Lower household income and higher population seems to correlate with higher poverty rates, which seems obvious, but this is countered somewhat by the very high household incomes that are found in very low and medium population areas. The very low population purple box is most likely Alexandria which goes against the overall trend shown of more population = more income = lower poverty. The collection of very high household income found in the medium population centers was pointed out in the last slide, but since there are more of those than the one Alexandria, it is more difficult to ignore and therefore more difficult to ignore why at a certain point very high population centers lose some correlation to household income.

The visualization of poverty rate by binned population and household income confirms the data shown in the previous visualizations. There is extreme high and low on the spectrum. There also seems to be a notable gap in the “High” and “Very Low” household income bins with each category only have one bin Future data could better analyze why those gaps exist.

**Geographic Distribution of Poverty Rates**

As with the first map, the population size is represented by the circle size. Color has been changed from being a legend identifier to now representing poverty rate with yellow being the highest poverty rates. Included is a hover message listing several data keys and values such as exact poverty rate, population, fort name, and the associated city. For this hover message, we have El Paso's Fort Bliss colored yellow and serving as one of the two extreme outliers of this dataset.

**CONCLUSION**

* The majority (19 of the 27) of the Army forts have poverty rates that fall above the national poverty line.
* There is a significant gap between the fort with the lowest poverty rate (Fort Belvoir) and the fort with the highest poverty rate (Fort Bliss).
* There is a [1] weak negative correlation (-0.39) between population size and poverty rate. As the population increases, the poverty rate should decrease.
* There is a [1] weak positive correlation (0.20) between population and household income. As the population increases, household income should increase.
* By limiting the study to the US Army branch only and to forts (large bases), it produced 27 bases. Limiting the study to one branch of the US military and limiting the size of US Army bases provided a glimpse into the possibilities. However, a study with all branches and all base sizes could provide greater insight. Further research would need to be done to determine what causations directly influence base location and poverty rate.
* The next iteration of this study should include socioeconomic factors that could create economically underserved communities around military bases, including light pollution, noise pollution (airplanes, helicopters, gunfire, etc.), and military personnel having a more transitory lifestyle.
* Different population data, such as gender, race, etc., should be considered.

[1] [researchgate.com](https://www.researchgate.net/figure/The-scale-of-Pearsons-Correlation-Coefficient_tbl1_345693737#:~:text=Upon%20examining%20the%20level%20of,a%20high%20correlation%20%5B42%5D%20.)