**Document link:** [**https://tinyurl.com/proj-see-the-girl**](https://tinyurl.com/proj-see-the-girl)

**DataDive master link:** [**https://datadive.datakind.org/**](https://datadive.datakind.org/)

**Executive summary**: The Delores Barr Weaver Policy Center works to ensure that all Floridians “See the Girl.” Through research, education, and advocacy programs, the Policy Center educates and advocates for girls, young women, and those identifying as female. The Policy Center relies on public data to do their work, and spends a significant amount of time in manual data processing and bespoke analysis and visualizations. In this project, you will help the Policy Center refine their data processing methods, create prototype visualizations for their future dashboards, and conduct data exploration that could support future research, policy and advocacy efforts.-

| **Requested skills**   * Documentation (no coding requirement) * Excel (macros, formulas, basic visualizations, documentation) * Tableau * Data visualization (non-Tableau) * Python or R for data processing * Python or R for data exploration |
| --- |

**Table of Contents**

# 

[**What Does DataDive Success Look Like?**](#_2pv58ur1318w) **3**

[Examples of success](#_2pnz2mihh35a) 3

[Example end-state visual outputs created by DataKind for other projects](#_z1r1phwt2qsz) 3

[**Introduction to the DataDive Project Partner**](#_ph4s24tnun2u) **4**

[**Problem**](#_d4iww6dwp7us) **5**

[Sample data](#_k1piv278ey2z) 5

[**Workstream #1: Data processing**](#_gr1olpdtng3p) **6**

[Data processing tasks](#_diyhcnxbkffa) 6

[**Workstream #2 Data Visualization**](#_ysn6mgrtzlh1) **10**

[Data Visualization Tasks](#_5co3leop0is5) 10

[**Workstream #3: Data Exploration**](#_59d9t68w1tj) **12**

[Data Exploration Tasks](#_i8qhdfj2el7e) 12

[**DataDive contact information**](#_mtw32yrthgts) **15**

[**DataDive volunteer resources**](#_fiht1vcp7qz9) **16**

[**Notes & ideas to share**](#_e9p5cd210fn5) **16**

[**DataDive volunteer sign-up**](#_yqco6kirqa6a) **17**

# 

# What Does DataDive Success Look Like?

*Success is about incremental learning and improvement*

**At DataDives, we seek to increase understanding of the issue area** we are partnered to work within - for our partners, for ourselves, and for the sector. We also seek to **turn these new understandings into additional insights so that our partners can better shape their own actions**. We provide catalyzing capacity, energy, and expertise to specific problems so that partners can take away knowledge that will help them take the next best step on their data science and data engineering journey.

| Examples of success   To increase understanding of this issue area, here are a few examples of what constitutes “success” for this project.   1. Conducting an exhaustive EDA on a shared dataset and being able to summarize what is - or is not - contained within that dataset 2. Demonstrating the application of multiple techniques to test the potential relationship between variables in a dataset and discussing the potential implications of those relationships 3. Creating prototype / proof of concept tools & visualizations that demonstrate the potential utility of new ways of visualizing and understanding data 4. Creating recommendations for how partner organizations might invest their resources as they continue on a data science journey 5. Showcasing the “Art of the Possible” by creating specific case studies or examples of how data science techniques can lead to better/faster/more insightful/less resource intensive outcomes 6. Designing the workflows that would allow for repeatable data extraction from multiple sources, combination, and visualization |
| --- |

## Example end-state visual outputs created by DataKind for other projects

Here are some examples of visual outputs that have been created at other DataKind events. **Note that it is unlikely that any one project or team will get all the way to these outputs in the course of the DataDive, and that’s okay!** These are example targets of insights and frameworks we’d like to provide in the future, and we’ll build on the work of the DataDive to get there - each of these relies on several streams of data and allows the end-user to more quickly drive decision-making.

* <https://www.datakind.org/blog/givingtuesday-report>
* <https://www.datakind.org/blog/unlocking-hidden-insights-in-the-social-sector-with-natural-language-processing>
* <https://www.datakind.org/projects/making-sense-of-text-data-to-help-disadvantaged-families>
* [https://www.datakind.org/blog/pivotal-for-good-with-crisis-text-li-ne-using-text-analytics-to-better-serve-at-risk-teens/](https://www.datakind.org/blog/pivotal-for-good-with-crisis-text-line-using-text-analytics-to-better-serve-at-risk-teens/)

# Introduction to the DataDive Project Partner

[The Delores Barr Weaver Policy Center](https://www.seethegirl.org/about-us/): born out of the Justice for Girls Reform Movement, the Policy Center opened in 2013 using political science and social science strategies that integrate the lived experiences of girls in the juvenile justice system. The work is grounded in research to inform the advocacy platform, training, and incubation of model programs to work with girls in each part of the juvenile justice system. The high caliber vision and leadership has positioned the organization for replication in other communities. Our goal is to ignite citizens to bring a unified voice to change laws, policies, and practices that contribute to inequitable treatment of girls that result in devastating long-term outcomes for girls, young women, and youth who identify as female. How do we do it? We work with girls as partners in the four cornerstones of our work: research, advocacy, training, and model programming.

* Example research that the Policy Center has published: [Notes to the Field: Girls and Secure Juvenile Detention *Barriers, Opportunities, and Recommendations*](https://www.evidentchange.org/sites/default/files/Notes%20to%20the%20Field%20Girls%20and%20Secure%20Juvenile%20Detnetion.pdf), [Girls in Secure Juvenile Detention in Florida](https://www.evidentchange.org/sites/default/files/Girls%20in%20Secure%20Juvenile%20Detention%20in%20Florida.pdf)
* Demonstrations of how the Policy Center uses data: [Sounding the Alarm](https://www.seethegirl.org/wp-content/uploads/2021/03/Sounding-the-Alarm_Full-report.pdf), [Status of Girls Well-being](https://www.seethegirl.org/wp-content/uploads/2019/10/WellBeingReportFinal.pdf)
* [Watch this quick welcome video that our partners made!](https://drive.google.com/drive/u/1/folders/19jiY56e2B7ojiKtBzLwYh0IP5WIliIpA)

# 

# Problem

The Delores Barr Weaver Policy Center (the Policy Center) exists to meet their mission statement: “We engage communities, organizations, and individuals through research, advocacy, training, and model programming to advance the rights of girls, young women, and youth who identify as female, especially those impacted by the justice system.” Successful engagements of all types - from research to programming - rely on the Policy Center accessing and using timely and trustworthy data on female-identifying individuals who are impacted by the Florida justice system. In order to obtain, assess, and describe the relevant data, the Policy Center team relies on a deeply resource-intensive and exclusively human-driven and manual process. The time costs of accessing the public data from multiple agencies, extracting that data, transforming it into standard data features, and loading it into a shared database are burdensome for a small and nimble team, limiting the frequency and versatility of analysis. If the team were to be able to produce similarly high quality engagement materials with a semi-automated process, they would remove a key resourcing bottleneck and be able to serve more individuals to better realize their mission statement.

## Sample data

* [Arrest and commitment for Black girls, 2018-19](https://drive.google.com/drive/folders/1Jw-JposQsor1BRHne5Ny6VHlngWrOdSr?usp=sharing)
* [Students by `race, 2015-2020](https://drive.google.com/drive/folders/1Fhzg6hHODMdGRX5ZIbokRkiFwdVrEHE1?usp=sharing)
* [School discipline data by race/ethnicity and gender, 2019-2020](https://drive.google.com/drive/folders/1E2LhSxoJcEl07uP9cOG-GZ4Shm1gOArf?usp=sharing)
* [Girls arrested, 2020](https://drive.google.com/drive/folders/1baTGdY4y-epCDw1MftZK2-or_hu-G0A-?usp=sharing)

# 

# Workstream #1: Data processing

* In the first step of their data-driven research process, the Policy Center team currently extracts data from a dozen-plus state-wide data repositories, each with its own extraction process. These data are extracted in multiple formats with multiple headers, measures, timeframes that the Policy Center team then cleans and stores for their analysis and visualization work.
* The Policy Center team has provided draft documentation on the data extraction processes, but would like to see this documentation expanded and improved upon, such that it can become a training tool for new staff. They are interested in the additional documentation of access processes for new-to-them datasets, such as the ACS (census) data.
* The Policy team has provided representative datasets that they have extracted to use for their work. In the second task of this workstream, the team would like to create or compile standard data dictionaries for the common datasets they use. The team accesses data as granular as an individual school and as aggregated as the entire state, but the most common unit of aggregation and analysis is the county. To prepare for data exploration and visualization tasks, the sample data should be compiled into a master dataset that contains variables aggregated to a county level.

## Data processing tasks

| **Tasks** | **Coding skills required** | **Claimed by**  **(insert your name)** | **Link to work** | **Additional notes** |
| --- | --- | --- | --- | --- |
| **Improve data extraction process documentation**  Using the [draft documentation](https://docs.google.com/document/d/183Fney6YPJh-Wus3Uk4b3wR39r-p-BhNE9UhOdHt7J4/edit), create a final draft of documentation that is standardized and easy to follow. Identify gaps in draft documentation coverage and fill in those gaps. | None | Jacqui  Jiselle  Jess Reeve  Roslyn (PDF Conversion) | Jess - https://drive.google.com/drive/folders/1Sz7K7Ubnz7s0Og-kuD2Un-a1CEzcSKuB | Jess - downloaded sample files from some sites listed in draft documentation (access through link). Suggested edits are comments in the draft documentation. |
| **Create/compile data dictionaries for sample data**  Many of the source data locations have data dictionaries, the information just needs to be compiled so that anyone looking at the sample data can understand the variables contained in the dataset. These data dictionaries can be stored in each sample data folder as an expansion of the “notes” document. Please link to any resources used. | None | Demetri | https://docs.google.com/spreadsheets/d/1BB\_RbwzB6nxEf-4bOX1TBNCvEeXV-lYikUmZNEFvWM8/edit?usp=sharing |  |
| **(subtask)** [Students by race, 2015-2020](https://drive.google.com/drive/folders/1Fhzg6hHODMdGRX5ZIbokRkiFwdVrEHE1?usp=sharing) |  | Roslyn  Heather-Cate |  | Finishing documentation for PDF conversion. PDF files downloaded on shared Google Drive  Heather-Cate: Added data dictionary to the Notes document |
| **(subtask)** [Girls arrested, 2020](https://drive.google.com/drive/folders/1baTGdY4y-epCDw1MftZK2-or_hu-G0A-?usp=sharing) |  | Srividhya Ammanur |  |  |
| **(subtask)** [Arrest and commitment for Black girls, 2018-19](https://drive.google.com/drive/folders/1Jw-JposQsor1BRHne5Ny6VHlngWrOdSr?usp=sharing) |  | Roslyn  Heather-Cate |  | Heather-Cate: Added data dictionary to the Notes document |
| **(subtask)** [School discipline data by race/ethnicity and gender, 2019-2020](https://drive.google.com/drive/folders/1E2LhSxoJcEl07uP9cOG-GZ4Shm1gOArf?usp=sharing) |  | Jennifer OU |  |  |
| **Clean and compile sample datasets at the Florida county level**  Create a master data file at the county level of the data contained across datasets | Minimal | Sarah Kelly  Baha Rababah | <https://drive.google.com/file/d/1bRY5jvphHJ9qXTShgnSYq1rCBt6NDma9/view?usp=sharing>  <https://docs.google.com/spreadsheets/d/1K927Go4_duErVxkjTO_dnYLGwbtftR4C/edit?usp=sharing&ouid=109657075124927861150&rtpof=true&sd=true>  <https://docs.google.com/document/d/1ajaxhFMSLx9UMZW6h1F1jwXntoff8LH7/edit?usp=sharing&ouid=109657075124927861150&rtpof=true&sd=true>  <https://drive.google.com/drive/u/0/folders/1cbafD4OmGx2gu6LbOgZMyJOifX6948HL> | Sarah Kelly- 3 of 4 cleaned and complied. Uploaded to google drive.  Baha Rababah cleaned the file Girls arrested 2020 |
| **Create a workflow for extracting PDF report data fr****om Tableau (**[**see notes here**](https://docs.google.com/document/d/1WIed6VxMYr7N6ZWUZPQyKt28C-89wP8z1igkbdHwG88/edit)**)**  Two reporting sources seem to have limited to data exports as PDF (vs CSV) - document the reporting process, save the raw PDFs, and provide your process for extracting  We’re looking for a process that would include approximately these 5 steps   1. Storage location 2. Naming convention for exports 3. Process for reading the PDF into R or Python for data extraction 4. Visualizing the new data 5. Cleaning the data as needed 6. Exporting and storing in a new data file or database 7. Appending that data to a master dataset |  | Brian Chu |  |  |
| **Extract additional datasets that might used by the Policy Center** [**and save here**](https://drive.google.com/drive/folders/16SMM3T0ZODw_L04deBqIVa6f9RrVwlcp?usp=sharing)  Referring to the [draft documentation for data extraction](https://docs.google.com/document/u/1/d/183Fney6YPJh-Wus3Uk4b3wR39r-p-BhNE9UhOdHt7J4/edit), you can see additional data sources not provided in the sample data. Additional data not currently that could be useful includes [ACS (census) data](https://drive.google.com/drive/folders/1NsZ2j10duCpXFL8rV9evoNxdpEZA6Vsc?usp=sharing), COVID-19 cases, unemployment, [housing loss](https://www.newamerica.org/future-land-housing/blog/explore-our-housing-loss-index/), [Stanford School Enrollment Project](https://purl.stanford.edu/zf406jp4427), among others. | None | Baha Rababah  Josie Wesseler | (Josie) I found an additional dataset on school discipline. Info on this sheet <https://docs.google.com/document/d/1QuJItcbm9Go-nVAp_N0TzbaOjtSioiXwQdIGUAOwjco/edit?usp=sharing> | PACT assessment data-- this helps us see the needs of girls, differences by gender and county on particular indicators of interest (sexual abuse history, mental health history, age at first arrest, etc) These are filtered under "risk factors" at the top of the screen. A useful visualization for us to share could be a table that summarizes all of these "risk factors" by gender for youth on probation, and then for youth on commitment to show the growing needs while deeper in the system. https://www.djj.state.fl.us/research/reports-and-data/interactive-data-reports/pact-profile/pact-profile-fy2017-18 |
| **(subtask)** Create or improve documentation for additional datasets |  |  |  |  |
| **(subtask)** Create/compile data dictionaries for additional datasets |  |  |  |  |
| **(subtask)** Clean and compile datasets at the Florida county level for additional datasets |  | Baha Rababah |  |  |
| *XXXX Additional task you propose* |  |  |  |  |
| *XXXX Additional task you propose* |  |  |  |  |
| *XXXX Additional task you propose* |  |  |  |  |
| *XXXX Additional task you propose* |  |  |  |  |
| *XXXX Additional task you propose* |  |  |  |  |

# 

# Workstream #2 Data Visualization

* The team conducts research-quality analytics for public consumption using multiple tools, but those are static and bespoke each time.
* There is an opportunity to create dynamic, interactive data visualizations and/or dashboards that would allow the Policy Center to share their analytics more broadly and engage additional community members in understanding and “seeing the girl.”
* Dashboards that the Policy Center has shared as inspiration
  + DJJ-DCF Dashboard (Florida Department of Juvenile Justice)
    - Pros: map/hover over county, filter ability, potential to merge race/ethnicity within gender.
  + [DMC-RED Profile FY2019-20](http://www.djj.state.fl.us/research/reports/reports-and-data/interactive-data-reports/disproportionate-minority-contact-reports/dmc-red-profile-fy2019-20) (Florida Department of Juvenile Justice)
  + [Florida Kids Count interface](https://www.aecf.org/interactive/databook?l=12) (the Annie E. Casey Foundation)
    - [Sample Kids Count reports](http://floridakidscount.org/images/2020-CWBI/2020_All_counties.pdf) (statewide aggregate and then by county for the selected variables).
    - Pros: nice snapshots also show change over time.

## Data Visualization Tasks

| **Tasks** | **Coding skills required** | **Claimed by (insert your name)** | **Link to work** | **Any additional notes** |
| --- | --- | --- | --- | --- |
| Create visualization in Tableau of information that the Policy Center compiles into [their snapshot documents](https://drive.google.com/drive/folders/1-GcsIENwT4OkQYh2L5yJQWueMXAguXFW?usp=sharing) | Minimal | Sam Sequeira  Shreyash Suryawanshi |  |  |
| Create tools and documentation to produce summary statistics from each sample dataset | Minimal - advanced, can be done in Tableau or any other analysis tool of choice |  |  |  |
| Create a [choropleth map](https://en.wikipedia.org/wiki/Choropleth_map) showcasing the relative distribution of key variables in the sample datasets across Florida | Minimal - advanced, can be done in Tableau or any other visualization tool of choice | Sukriti Chaudhary  Catharina Oliveira |  |  |
| **(subtask)** [Students by race, 2015-2020](https://drive.google.com/drive/folders/1Fhzg6hHODMdGRX5ZIbokRkiFwdVrEHE1?usp=sharing) |  | Pooja Purushothaman  Jerrie Ouyang | <https://docs.google.com/document/d/14llnrb6uNNHNpAQHHG6OKNvtIMaIcm0R1zyysX3qMlk/edit?usp=sharing>[Documentation]  <https://docs.google.com/spreadsheets/d/13mPCXa74RcsPTdqmDLa3Ve_FVAA5l-aN_2OhDFo4-EQ/edit?usp=sharing>[Final File] | Worked on the data set and converted into a more usable format  Jerrie& Pooja are interested in creating visualization post the event. |
| **(subtask)** [Girls arrested, 2020](https://drive.google.com/drive/folders/1baTGdY4y-epCDw1MftZK2-or_hu-G0A-?usp=sharing) |  | Cassie Zhang  Rima | https://public.tableau.com/app/profile/rima.hinnawi/viz/GirlsArrestedin2020/Dashboard1?publish=yes | Worked on this task and uploaded on tableau public--Rima |
| **(subtask)** [Arrest and commitment for Black girls, 2018-19](https://drive.google.com/drive/folders/1Jw-JposQsor1BRHne5Ny6VHlngWrOdSr?usp=sharing) |  | Jason Freeman |  |  |
| **(subtask)** [School discipline data by race/ethnicity and gender, 2019-2020](https://drive.google.com/drive/folders/1E2LhSxoJcEl07uP9cOG-GZ4Shm1gOArf?usp=sharing) |  | Kristine Rosales  Josh Ash  Josie Wesseler | (Josie) Limited visualization and an additional dataset for future work: <https://docs.google.com/document/d/1QuJItcbm9Go-nVAp_N0TzbaOjtSioiXwQdIGUAOwjco/edit?usp=sharing> |  |
| *XXXX Additional task you propose* |  |  |  |  |
| *XXXX Additional task you propose* |  |  |  |  |
| *XXXX Additional task you propose* |  |  |  |  |
| *XXXX Additional task you propose* |  |  |  |  |
| *XXXX Additional task you propose* |  |  |  |  |

# Workstream #3: Data Exploration

* The Policy Center is eager to explore the potential of analytics and to see the possibilities that data science might offer in creating an understanding of the status of girls in Florida and the need for recommended action. This workstream is intended to be exploratory and to demonstrate the possibility (or lack thereof) of deriving additional insights as to the long-term outcomes for girls, young women, and youth who identify as female.
* Datasets that might be of interest to compare with outcomes in the sample datasets include: [ACS (census) data](https://drive.google.com/drive/folders/1NsZ2j10duCpXFL8rV9evoNxdpEZA6Vsc?usp=sharing), COVID-19 cases, data from the [City of Orlando’s Open Data Portal](https://data.cityoforlando.net/) (and other city or county data portals), [public health data](https://guides.uflib.ufl.edu/precisionpublichealth/datalibrary), county-level employment/unemployment data, [housing loss](https://www.newamerica.org/future-land-housing/blog/explore-our-housing-loss-index/), [Stanford School Enrollment Project](https://purl.stanford.edu/zf406jp4427), [USDA ERS resource (county level poverty, population, unemployment, education)](https://www.ers.usda.gov/data-products/county-level-data-sets/), among others.
* Example data exploration papers that might be helpful as starting points:
  + <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7456591/>
  + <https://www.ojp.gov/pdffiles1/ojjdp/226358.pdf>
  + <https://www.equalmeasure.org/wp-content/uploads/2019/12/EquityCounts_OYF-Measures-Brief_EqM_FINAL_Oct19.pdf>
  + <https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2720072>
* Exploration question ideas
  + The Policy Center is actively looking to improve the lives of girls, particularly those who have interactions with the justice system, so it might be interesting to understand if there are relationships between arrest or discipline outcomes and any socioeconomic, health, demographic or other variables
  + The Policy Center is interested in understanding outcomes for girls, it might be interesting to look at if outcomes cluster together - e.g. do counties that have higher arrest rates also have higher school discipline rates?
  + Deep dives into particular geographies
* From The Policy Center
  + What are girls committed for and how does it differ from the boys? We usually break this down in pie charts to show felony, misdemeanor, and zero in on the "other offenses" since girls are more likely to be incarcerated/committed for non-law technical violations of probation as a proportion of total offenses than boys (1/5 compared to 1/10).
  + What counties have the highest incarceration rates, greatest disparities for Black girls?
  + How are girls in XYZ community doing- how many are being suspended, arrested in school, arrested in the community, receiving civil citation (and being diverted)- are the trends improving?
  + how many girls are ending up in the system for domestic violence related offenses when they would've been eligible for a civil citation (but instead charged as a battery). This data is available on the civil citation dashboard when you filter for domestic violence--but unfortunately it is limited to only those that were eligible, we don't know how many of the total arrests of girls are for family related offenses because we cannot disaggregate battery charge any further.

## Data Exploration Tasks

| **Tasks** | **Coding skills required** | **Claimed by (insert your name)** | **Link to work** | **Any additional notes (e.g. type of analysis)** |
| --- | --- | --- | --- | --- |
| Compute [Relative Rate Index (RRI)](https://www.multco.us/dcj/racial-ethnic-disparities-relative-rate-index) for sample datasets  From the Policy Center: *this could be a good way to quickly “see” overrepresentation for Black girls by point in system* | Minimal - advanced, can complete in analytics tools of choice | Ojaswini Pandey  Anuoluwapo |  | Relative Rate index that FL DJJ uses and the visualizations: <https://www.djj.state.fl.us/research/reports-and-data/interactive-data-reports/disproportionate-minority-contact-reports/dmc-red-profile-fy2019-20> |
| **(subtask)** [Students by race, 2015-2020](https://drive.google.com/drive/folders/1Fhzg6hHODMdGRX5ZIbokRkiFwdVrEHE1?usp=sharing) |  | Kristine Rosales  Sai Pravallika Myneni  Nur Zhan | https://colab.research.google.com/drive/17EPsu8A\_\_2zZ3a1QqwZFVpexNaXPyq4U?usp=sharing |  |
| **(subtask)** [School discipline data by race/ethnicity and gender, 2019-2020](https://drive.google.com/drive/folders/1E2LhSxoJcEl07uP9cOG-GZ4Shm1gOArf?usp=sharing) |  | Rahul Paleja  Kristine Rosales  Nur Zhan  Ojaswini Pandey | https://github.com/nulyaka/DataKind/blob/main/Basic%20EDA.ipynb |  |
| **(subtask)** [Girls arrested, 2020](https://drive.google.com/drive/folders/1baTGdY4y-epCDw1MftZK2-or_hu-G0A-?usp=sharing) |  | Rahul Paleja  Brian Chu  Ojaswini Pandey  Sai Pravallika Myneni  Anuoluwapo |  |  |
| Conduct EDA on sample data in addition to socio-economic, health, and other variables.  *(note* [*task in Workstream 1*](#eti0l6olidt5) *for compiled datasets - they might be useful here, if you compile datasets,* [*please document and share*](https://drive.google.com/drive/folders/16SMM3T0ZODw_L04deBqIVa6f9RrVwlcp?usp=sharing)*!)* | Minimal - advanced, can complete in analytics tools of choice | Dorna  Suman Ghosh |  | Look at trends that may explain the highlighted counties in ‘Arrest and Commitment 2018’  Suman: Was exploring the data in girls arrested 2020 excel. Jupyter notebook uploaded in data folder. |
| Conduct EDA on sample data in addition to socio-economic, health, and other variables.  *(note* [*task in Workstream 1*](#eti0l6olidt5) *for compiled datasets - they might be useful here, if you compile datasets,* [*please document and share*](https://drive.google.com/drive/folders/16SMM3T0ZODw_L04deBqIVa6f9RrVwlcp?usp=sharing)*!)* | Minimal - advanced, can complete in analytics tools of choice | Ruchika |  |  |
| Conduct EDA on sample data in addition to socio-economic, health, and other variables.  *(note* [*task in Workstream 1*](#eti0l6olidt5) *for compiled datasets - they might be useful here, if you compile datasets,* [*please document and share*](https://drive.google.com/drive/folders/16SMM3T0ZODw_L04deBqIVa6f9RrVwlcp?usp=sharing)*!)* | Minimal - advanced, can complete in analytics tools of choice | Mark Zollner | <https://drive.google.com/drive/u/2/folders/14yOuL70dhQ4WmVjDSpw5yUGlJlF9xjMG> | The county names were not consistent between the poverty level dataset and the arrest excel file so I manually cleaned them up to save time. Ideally this should have been automated. |
| Conduct EDA on sample data in addition to socio-economic, health, and other variables.  *(note* [*task in Workstream 1*](#eti0l6olidt5) *for compiled datasets - they might be useful here, if you compile datasets,* [*please document and share*](https://drive.google.com/drive/folders/16SMM3T0ZODw_L04deBqIVa6f9RrVwlcp?usp=sharing)*!)* | Minimal - advanced, can complete in analytics tools of choice | Farah |  |  |
| Conduct EDA on sample data in addition to socio-economic, health, and other variables.  *(note* [*task in Workstream 1*](#eti0l6olidt5) *for compiled datasets - they might be useful here, if you compile datasets,* [*please document and share*](https://drive.google.com/drive/folders/16SMM3T0ZODw_L04deBqIVa6f9RrVwlcp?usp=sharing)*!)* | Minimal - advanced, can complete in analytics tools of choice |  |  |  |
| *XXXX Additional task you propose* |  |  |  |  |
| *XXXX Additional task you propose* |  |  |  |  |
| *XXXX Additional task you propose* |  |  |  |  |
| *XXXX Additional task you propose* |  |  |  |  |
| *XXXX Additional task you propose* |  |  |  |  |

# DataDive contact information

Project Support: Caitlin Augustin ([caitlin@datakind.org](mailto:caitlin@datakind.org))

Technical support: Sophia ([sophia@datakind.org](mailto:sophia@datakind.org)), Emily Yelverton ([emily@datakind.org](mailto:emily@datakind.org))

# DataDive volunteer resources

1. Slack workspace: <https://tinyurl.com/September-DataDive>
2. Slack channel: #proj-see-the-girl
3. Dedicated zoom: <https://datakind-org.zoom.us/my/kgjohnson?pwd=LzVmblJYUGNxWVh1aVRTRmprYStXdz09>
4. Additional resources for small-group collaboration
   1. [Slack calls](https://slack.com/help/articles/216771908-Make-calls-in-Slack) - you can call people 1-to-1 directly in slack
   2. [Google meet](https://support.google.com/meet/answer/9302870?co=GENIE.Platform%3DDesktop&hl=en) - you can schedule a meeting for now, or for later in Google meet if you have a google email account
5. Github: <https://github.com/datakind/Sept21-see-the-girl>
6. Questions for partners: <https://docs.google.com/document/d/1jWCJEg5ijBJHBBLINPilAYjwsZZLhRaVTsuUjICnHFI/edit>
7. Volunteer Materials Google Drive: https://drive.google.com/drive/u/1/folders/1cbafD4OmGx2gu6LbOgZMyJOifX6948HL

# 

# Notes & ideas to share

# 

# values from the ACS data https://www.census.gov/data/developers/data-sets/acs-1year/notes-on-acs-estimate-and-annotation-values.html

# At the NCES (National Center for Education Statistics) website they have a new visualization tool which visualizes a few of the asks .

<https://go.usa.gov/xMkZG>

It represents in the linked view demographic race in school concentrations. It has gender also. I am working on to link this data with the crime and incidence data as well. I will see what I can do. [*School Survey on Crime and Safety (SSOCS)*](https://nces.ed.gov/surveys/ssocs/index.asp) found at <https://nces.ed.gov/surveys/ssocs/index.asp?FType=2> nces.ed.gov**nces.ed.gov** [**School Survey on Crime and Safety (SSOCS) - Overview**](https://nces.ed.gov/surveys/ssocs/index.asp)nces.ed.gov**nces.ed.gov** [**School Survey on Crime and Safety (SSOCS) - Overview**](https://nces.ed.gov/surveys/ssocs/index.asp?FType=2)

# DataDive volunteer sign-up

Fill this out - if you have a Github account and actively use it, we’d love to add you to the repository for this project. If you don’t want to use Github, please upload your project work into [this shared Google Drive folder](https://drive.google.com/drive/folders/1cbafD4OmGx2gu6LbOgZMyJOifX6948HL?usp=sharing).

| **Name** | **Email address** | **Github (if you have it)** |
| --- | --- | --- |
| Caitlin Augustin | caitlin@datakind.org | augustincaitlin |
| Sukriti Chaudhary | sukritichaudhary30@gmail.com | sukritichaudhary30 |
| Jiselle Jones | jisellejones@gmail.com | jisellejones |
| Sarah Kelly | sarahkelly@tabsanalytics.com |  |
| Baha Rababah | baharababah@yahoo.com | baharababah |
| Shreyash Suryawanshi | shreyashvs96@gmail.com | shreyashvs96 |
| Bill Dwyer |  | BillDwyer |
| Rahul Paleja | rrpaleja@gmail.com | rrpaleja |
| Hong Yee GAN | osofunnie@gmail.com | HongYeeGAN |
| Jacqui Georgi | jhgeorgi@gmail.com | jhgeorgi |
| Brian McMurray | thebrianmcmurray@gmail.com | thebrianmcmurray |
| Roslyn Hurt-Steverson | RoslynSteverson@gmail.com | RosSmiles |
| Demetri Lee | demetri.lee@gmail.com | demetri77 |
| Catharina Oliveira | catharinafoliveira@gmail.com | cathfoliveira |
| Sam Sequeira | ssequeira22@gmail.com | ssequeira22 |
| Ruchika Sah | ruchika.sah@gmail.com | ruchika817 |
| Cassie Zhang | cassie202107@gmail.com | water2015sit |
| Ojaswini Pandey | ojaswinipandey@gmail.com | OjaswiniPandey |
| Sai Pravallika Myneni | mynenisp1703@gmail.com | Pravallika-Myneni |
| Brian Chu | brkchu@gmail.com | bchugit |
| Farah | isthatso@gmail.com | 3fa |
| Anuoluwapo Balogun | [Ifeoluwapobalogun1@gmail.com](mailto:Ifeoluwapobalogun1@gmail.com) | Designegycreatives |
| Jess Reeve | [jessicareeve@tabsanalytics.com](mailto:jessicareeve@tabsanalytics.com) |  |
| Kristine Rosales | profkrisrosales@gmail.com | krosales21 |
| Thirumal Sethusivaram | thirumal1995@gmail.com |  |
| Nur Zhan | nurlanzhangali@gmail.com | nulyaka |
| Malika Mohan | malikamohan99@gmail.com | malikamohan01 |
| rima | rima@restonyoga.com | rimahinn |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |