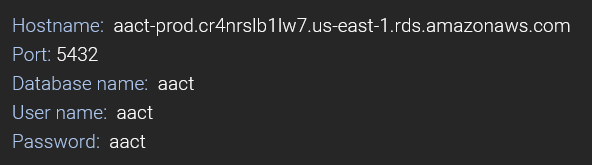
**DS 745: Project 1: Visualization Project**

Student: Connie Sosa Project due date: Feb. 19, 2017

* **What is the source of your data?**

The source of the data is [AACT](http://aact.ctti-clinicaltrials.org/) (Aggregate Analysis of ClinicalTrials.gov) database, which is sponsored by [Clinical Trials Transformation Initiative](https://www.ctti-clinicaltrials.org/) (CTTI). CCTI is a public-private partnership, its members are comprised of government agencies, industry representatives, patient advocacy groups, professional societies, academic institutions, and other interested parties. This database provides an analyzable dataset of all study information contained in ClinicalTrials.gov.

AACT is a cloud-hosted publicly available PostgreSQL relational database. I’ve connected to the AACT database from both Tableau and R/RStudio, evaluated both software tools, and chose Tableau as the tool of choice for this project. Please use the following credentials to [connect](http://aact.ctti-clinicaltrials.org/connect) to PostgreSQL Server from Tableau Public.



* **What is your visualization's purpose?**

The purpose of this visualization is to identify popular types of treatment interventions being studied for gestational diabetes patients. Determine the number of participants for intervention such as medical strategy, treatment, or device for each stage of this clinical research. With this information, one can further examine and make actionable recommendation for medical approach that works best for patients with gestational diabetes in future analytical visualization.

* **What does your visualization show; i.e. what is your portrayed visualization “story”?**

This visualization shows that treatment intervention ‘Device’ has the most number of participants for the clinical studies. It has about four times as many enrollees compare with the next popular intervention type studied. It further shows that Phase III has the most participants for this research. In this phase, evidence suggesting effectiveness of the drug has been obtained, and it’s the last phase before FDA approves the treatment under research.

* **What are your axes?**

The x-axis shows the five possible types of treatment intervention for this clinical trial studies. The treatment intervention types were Behavioral, Device, Dietary Supplement, Drug, and Procedure. The y-axis shows the total number of enrollment. The enrollment number ranges from 0 to 75,000. Additional variables were added to the different iterations of the visualization.

* **What variables and records (observations) you are using?**

The following variables were used to create the visualization.

Phase from the studies table. It’s a category field, phases I through IV and early phase I were used. Please note, phase I/II and phase II/III were also in the studies, but were excluded in this visualization.

Intervention Type from the interventions table. It’s a category field, the possible categories for the visualization of gestational diabetes studies were Behavioral, Device, Dietary Supplement, Drug, and Procedure.

|  |  |  |
| --- | --- | --- |
| **Intervention Type** | **Example** | **Note** |
| Behavioral | lifestyle counseling |  |
| Device | glucose monitor | including sham (placebo device) |
| Dietary Supplement | vitamins, minerals |  |
| Drug |  | including placebo |
| Procedure |  |  |

The records (observations) used were Enrollment from the studies table. It’s a metrics that measures the total number of participants that were enrolled in the studies.

* **Was any cleaning required? If so, what was it?**

There wasn’t any cleaning required, but the data were filtered with the following criteria. Only interventional studies were used, observational studies were not included in this visualization. Out of 16 possible diabetes related conditions, only gestational diabetes studies were included in this visualization. Phase 0 (early phase I), I, II, III, and IV were included, phase I/II and phase II/III were excluded in this visual.

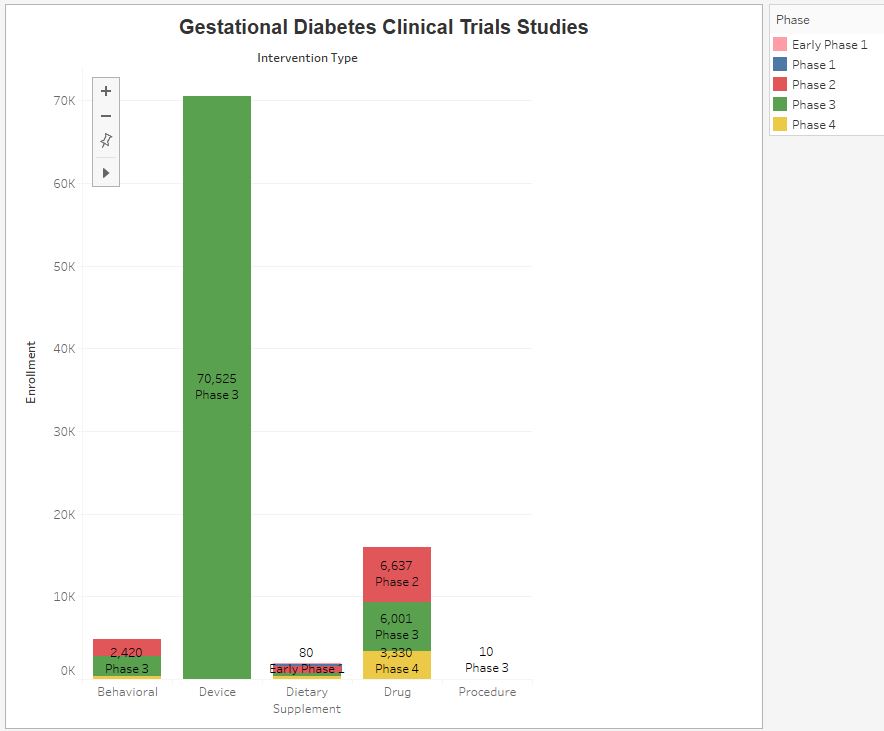
* **Develop three iterations for your visualization story, iteratively improving upon the prior version. Discuss pros and cons for each iteration.**

Please see Tableau file Project1Sosa.twb for the three iterations of the visualization story.

***Iteration 1:***

**Pros:** This stacked bar chart nicely shows the five different intervention type for the gestational diabetes clinical research and the color coded breakdown for each of the study phase. This is great to get an overall view of the relative number for each intervention type studied.

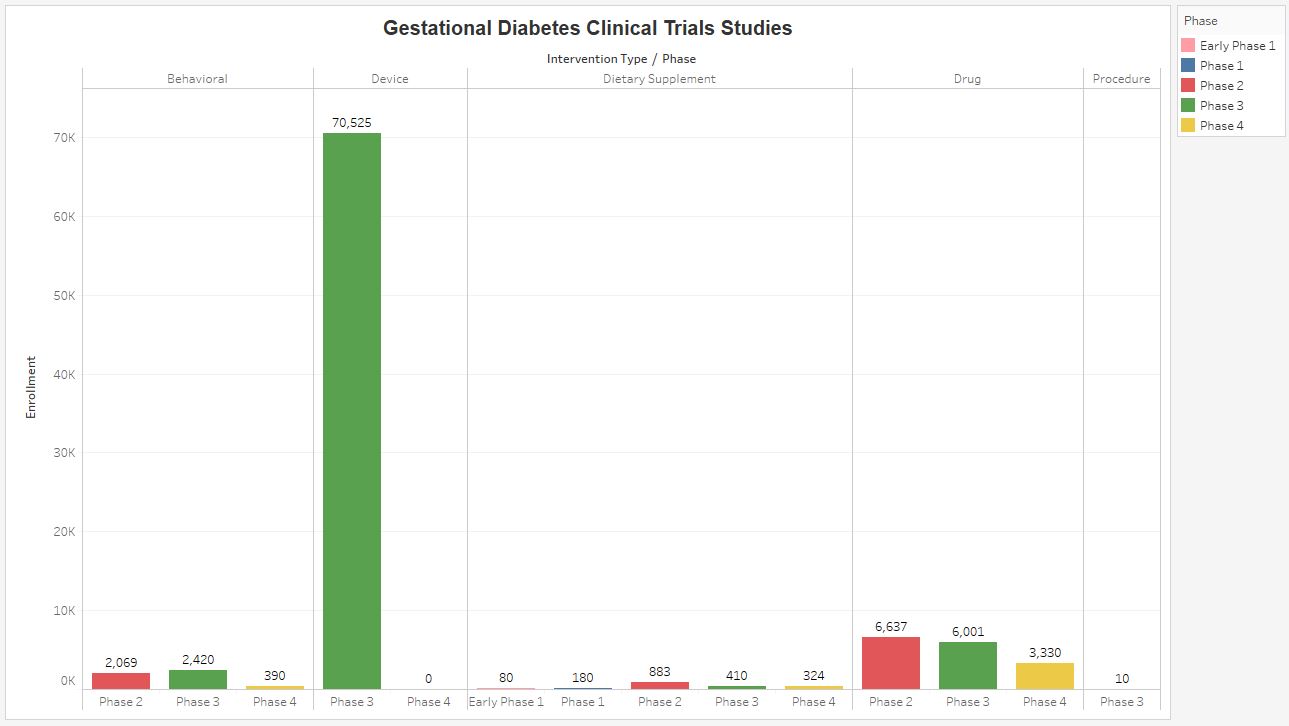
**Cons:** It is difficult to see the enrollment number of phases with a small number of enrollment. Additionally, labeling of small segments clutter the chart and make it difficult to identify the associated enrollment number corresponds to that phase. Also, it’s difficult to make comparison of the enrollment number for different phases of the studies.



***Iteration 2:***

**Pros:** This side-by-side bar chart shows the five different intervention type for the gestational diabetes clinical research and it segmented the phase by different colors. With this iteration, the enrollment number is clearly shown and easily seen at the top of each bar. Labels are no longer cluttered.

**Cons:** The graph is now wider, and is not within the eye span. Scrolling might be necessary depends on the size of the tool window. The fact that not all five phases are present and the resulting variable width of different intervention type makes pattern-finding difficult, which makes comparisons of different phases challenging.

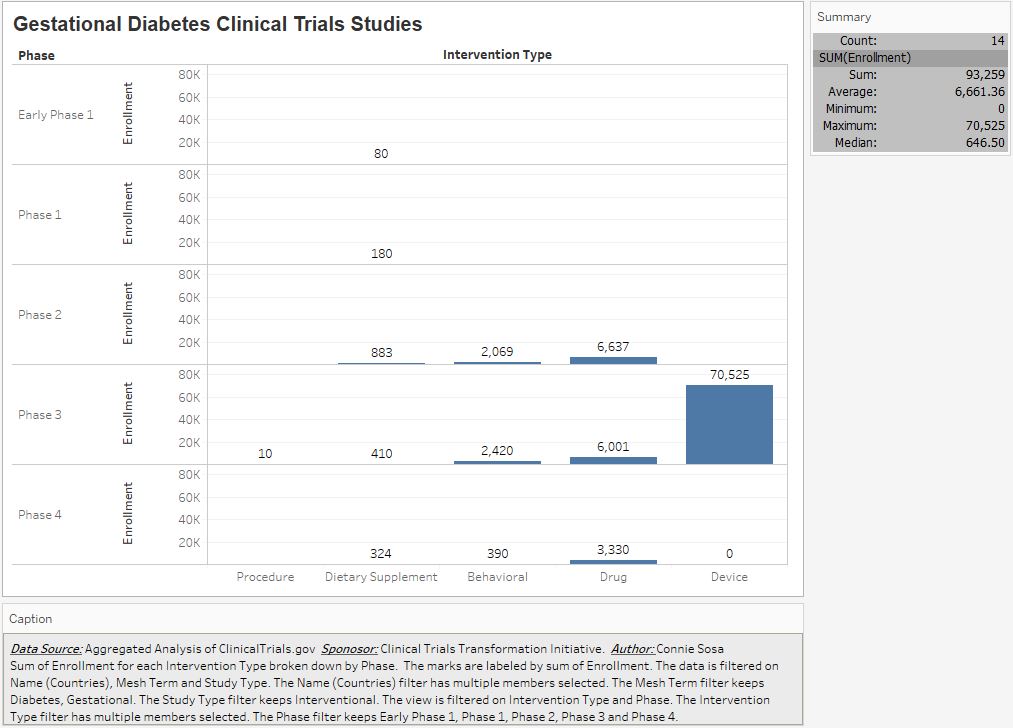


***Iteration 3:***

**Pros:** To properly describe the evidence as recommended by Tufte, data source, sponsor, author, and relevant issues were added to this visualization. Use of varying colors to segment different phase of the studies were removed. Rather, phase dimension was moved to the y-axis to give a better overall visual of the multivariate nature of this analysis. With multiple vertical bar charts, each represents the number of enrollments by intervention type, and were divided into five different phases for the clinical studies. This type of chart is helpful since large differences observed between the different intervention types.

The presence of a disproportionately large enrollment number for the particular intervention type made the smaller numbers difficult to see in other type of charts. The visualization makes it relatively easy to make enrollment comparisons of different intervention type and different phase.

**Cons:** This visualization makes the comparison of the sum total amount of enrollment for different intervention type is more difficult. By adding another vertical bar chart that shows the total enrollment for all five phases might correct this deficiency.



**References:**

http://www.dcri.org/

https://prsinfo.clinicaltrials.gov/definitions.html#InterventionType

https://www.nlm.nih.gov/services/ctphases.html

http://aact.ctti-clinicaltrials.org/

https://www.ctti-clinicaltrials.org/

|  |  |
| --- | --- |
| **Phase** | **Description** |
| Early Phase 1 | Formerly listed as ‘Phase 0’ |
| Phase 1 | Includes initial studies to determine the metabolism and pharmacologic actions of drugs in humans, the side effects associated with increasing doses, and to gain early evidence of effectiveness; may include healthy participants and/or patients. |
| Phase 2 | Includes controlled clinical studies conducted to evaluate the effectiveness of the drug for a particular indication or indications in participants with the disease or condition under study and to determine the common short-term side effects and risks. |
| Phase 3 | Includes trials conducted after preliminary evidence suggesting effectiveness of the drug has been obtained, and are intended to gather additional information to evaluate the overall benefit-risk relationship of the drug. |
| Phase 4 | Studies of FDA-approved drugs to delineate additional information including the drug's risks, benefits, and optimal use. |