

|  |
| --- |
| Arden University |
| MSc Data Analytics and Information Systems Management |
| [Data Visualisation and Interpretation (DAT7002BNM)](https://moodle.bl.rdi.co.uk/course/view.php?id=3309) |
| Portfolio: Data Visualisation and Interpretation |
| STU206280 |
| Ms. Zingsho Vashum |
| 2432 Words |

**Table of Contents**

[1 Introduction 3](#_Toc171938319)

[2 Data Overview for King Crab Dataset 3](#_Toc171938320)

[3 Data Cleaning and Preprocessing 4](#_Toc171938321)

[3.1. Loading Data: 5](#_Toc171938322)

[3.2 Data cleaning 6](#_Toc171938323)

[3.3 Data Integration: 8](#_Toc171938324)

[3.4 Data Transformation: 9](#_Toc171938325)

[3.5 Data Quality Checks: 9](#_Toc171938326)

[4 Data Visualization and Analysis 11](#_Toc171938327)

[**Data Visualization in Tableau** 11](#_Toc171938328)

[4.1 Bar Charts 11](#_Toc171938329)

[4.2 Line Charts 14](#_Toc171938330)

[4.3 Scatter Plots 15](#_Toc171938331)

[4.4 Box and Whisker Plots 16](#_Toc171938332)

[4.5 Heat Maps 18](#_Toc171938333)

[4.6 Text label 20](#_Toc171938334)

[5 Storytelling with Data 21](#_Toc171938335)

[5.1. Implications and Insights 21](#_Toc171938336)

[6 Conclusion and Recommendations 22](#_Toc171938337)

[7 References 23](#_Toc171938338)

**1 Introduction**

Visualizations are essential tools in data analysis, facilitating understanding and decision-making processes (Batt et al., 2020). This report delves into the health status of King Crabs in Alaska's coastal waters, specifically focusing on males (M), females (F), and infants (I) of the Red King Crab species (Paralithodes camtschaticus). These crabs are prized for their economic value and play a crucial role as keystone species in marine ecosystems, particularly in the Bering Sea and along Alaska's coast (NOAA Fisheries - Red King Crab).

King Crabs face significant ecological challenges, historically evidenced by the collapse of major stocks, such as those in Bristol Bay in the early 1980s, leading to the closure of commercial fishing areas (Westphal, 2011). The decline has been attributed to factors like overfishing, habitat degradation, and climate variability impacts (Stevens & Jewett, 2014).

Recent research efforts have focused on understanding the complex life cycle and ecological interactions of King Crabs. These studies investigate factors influencing population health, including reproduction rates, growth patterns, and responses to environmental stressors (Westphal, 2011). Such insights are crucial for formulating effective conservation and management strategies to sustain King Crab populations.

This report utilizes Tableau for visual data analysis, aiming to provide clear insights into key parameters such as growth rates, shell conditions, and mortality rates across different age groups of King Crabs. By presenting analytically robust visualizations, this study seeks to inform stakeholders, including government officials and fisheries management bodies, to support evidence-based decision-making for conservation and sustainable fisheries management in Alaska.

In summary, assessing the health of Alaska's King Crab populations involves integrating ecological and economic perspectives. This report underscores the importance of visual data analysis in generating meaningful insights into King Crab dynamics, thereby contributing to the preservation of this vital marine resource and supporting the sustainable management of Alaska's fisheries.

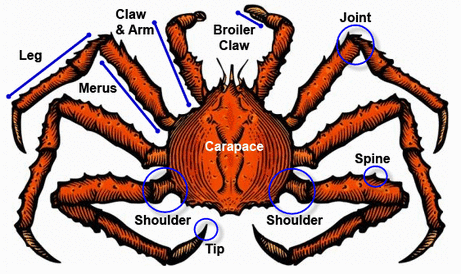


Fig.1. King Crab Alaska

### 

### 2 Data Overview for King Crab Dataset

**Introduction**

This report provides an overview of the King Crab dataset, detailing its structure, preprocessing steps, and potential insights. The dataset includes measurements and weights of King Crabs collected over multiple years, offering valuable information for biological and environmental studies.

**Dataset Description**

- **Dataset Name:** King Crab Dataset

- **Data Source:** Excel file provided

- **Data Period:** Contains data for multiple years (Lobster 2018 and Lobster 2019)

**Data Structure**

The dataset comprises the following columns:

1. **Sex:** Gender of the crabs, categorized as male (M), female (F), and potentially infants (I).

2. **Length(mm):** Measurement of the crab's length in millimeters.

3. **Diameter(mm):** Measurement of the crab's diameter in millimeters.

4. **Height(mm):** Measurement of the crab's height in millimeters.

5. **WholeWeight(g):** Total weight of the crab in grams.

6. **ShuckedWeight(g):** Weight of the crab's meat (edible portion) in grams.

7. **SellWeight(g):** Weight of the crab's shell after shucking in grams.

8. **Year:** The year in which the data was recorded.

### 3 Data Cleaning and Preprocessing

The dataset underwent several preprocessing steps using pandas to ensure accuracy and reliability for analysis:

## **3.1. Loading Data:**

* The dataset was loaded from an Excel file using pandas.

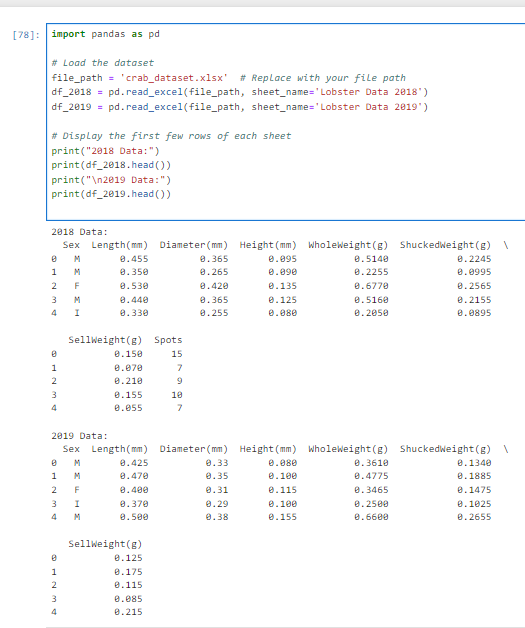
****

Fig.3.1. Loading Data

**Output Explanation**:

* + **2018 Data**: This section shows the first few rows of the 'Lobster Data 2018' sheet. Each row represents a crab with attributes like Sex, Length(mm), Diameter(mm), Height(mm), WholeWeight(g), ShuckedWeight(g), SellWeight(g), and Spots. For example, the first row indicates a male crab (Sex = 'M') with specific measurements and weights.
  + **2019 Data**: Similarly, this section shows the first few rows of the 'Lobster Data 2019' sheet. The structure is the same as for 2018, but it contains data for a different year.

**3.2 Data cleaning**

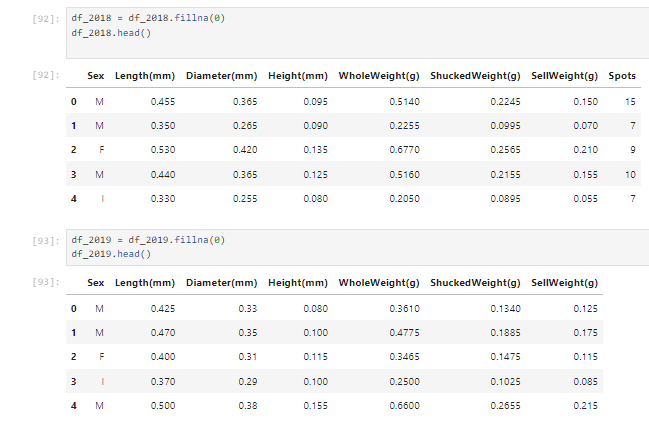
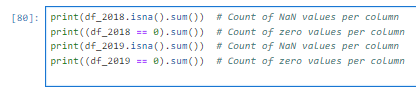
****

Fig.3.2.1 Data cleaning

**fillna(0)**:

* df\_2018 = df\_2018.fillna(0) and df\_2019 = df\_2019.fillna(0) fill any missing (NaN) values in your DataFrames with zero (0).
* This operation replaces all NaN values in the entire DataFrame with zeros.

Analyzing the presence of missing values (NaN) and zero values in your df\_2018 and df\_2019 DataFrames.



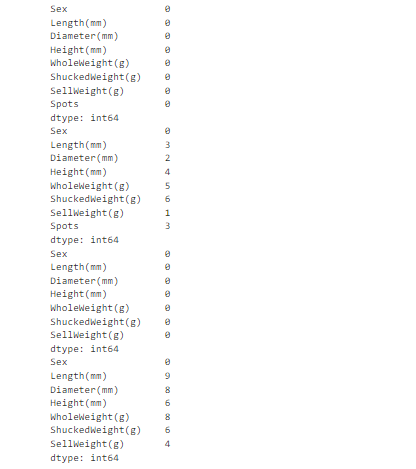


Fig.3.2.2 Count the NaN and Zeros

Replace zero values with mean of the field if they are meant to indicate missing data. (Rubin, D. B. (1976)).

Missing values were identified and addressed using appropriate techniques such as imputation or removal.



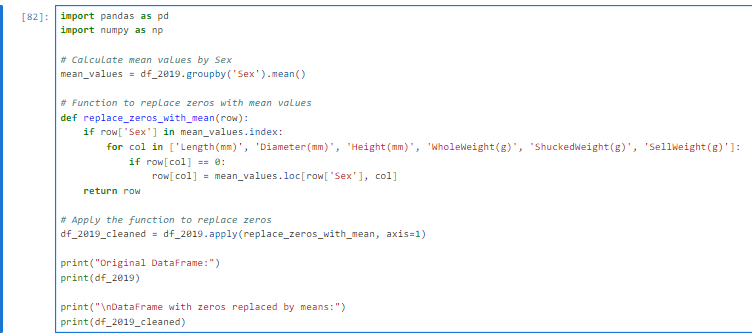


Fig.3.2.3 Replacing Zero with Mean

Calculate the mean values for each numeric column (Length(mm), Diameter(mm), Height(mm), WholeWeight(g), ShuckedWeight(g), SellWeight(g)) grouped by the 'Sex' column. This allows you to later replace zeros with these mean values specific to each sex group.

## **3.3 Data Integration:**

- Data from multiple sheets (2018 and 2019) were combined into a single DataFrame for comprehensive analysis.( Kimball, R., & Caserta, J. (2004)).

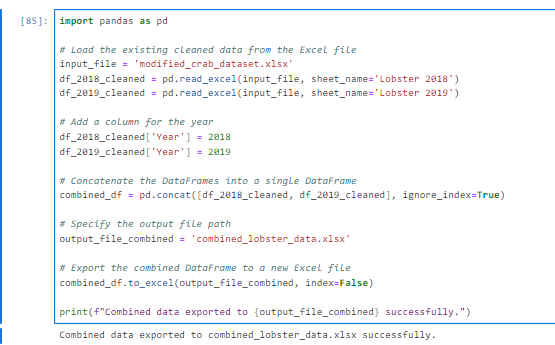


Fig.3.3 Data Integration

Adding a new column named 'Year' to each DataFrame (df\_2018\_cleaned and df\_2019\_cleaned) with the respective year values (2018 and 2019). And then concatenated both 2018 and 2019 sheets together for the better analysis. Exports the consolidated data to a new Excel file.

## **3.4 Data Transformation:**

- Data was converted into appropriate formats (e.g., numerical values, date formats) to facilitate

## **3.5 Data Quality Checks:**

- Ensured consistency and accuracy by resolving inconsistencies and removing duplicates.

**Analysis**

1. Descriptive Statistics:

- Basic statistics (mean, median, standard deviation) were calculated for numerical attributes to understand central tendencies and variability.

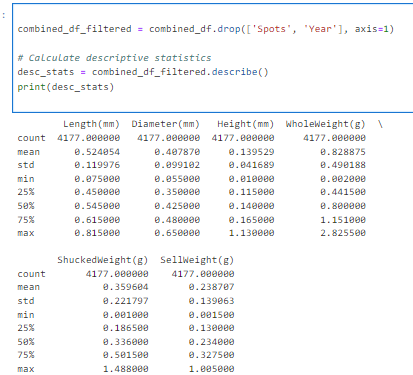


Fig.3.5 Descriptive Statistics

Insights from the descriptive statistics

* **Length(mm)**: The average length of the crabs is approximately 0.524 mm, with a standard deviation of 0.120 mm. The shortest crab observed was 0.075 mm, and the longest was 0.815 mm.
* **Diameter(mm)**: The average diameter is around 0.408 mm, with a standard deviation of 0.099 mm. The smallest diameter observed was 0.055 mm, and the largest was 0.650 mm.
* **Height(mm)**: The average height is about 0.140 mm, with a standard deviation of 0.042 mm. The smallest height observed was 0.010 mm, and the largest was 1.130 mm.
* **WholeWeight(g)**: The average whole weight of the crabs is approximately 0.829 grams, with a standard deviation of 0.490 grams. The lightest crab weighed 0.002 grams, and the heaviest weighed 2.826 grams.
* **ShuckedWeight(g)**: The average weight of the meat (shucked weight) is about 0.360 grams, with a standard deviation of 0.222 grams. The minimum observed shucked weight was 0.001 gram, and the maximum was 1.488 grams.
* **SellWeight(g)**: The average weight of the crab meat sold is approximately 0.239 grams, with a standard deviation of 0.139 grams. The smallest sell weight observed was 0.002 grams, and the largest was 1.005 grams.

**Conclusion**

The King Crab dataset provides valuable insights into the physical characteristics and weights of crabs over time and by gender. This overview offers a foundational understanding of the dataset structure and the analytical approaches applied to derive meaningful insights.

### 4 Data Visualization and Analysis

**Introduction**

Data visualization and analysis are crucial for deriving meaningful insights from the King Crab dataset. This section details the process of visualizing and analyzing the data to uncover patterns, trends, and relationships among various attributes. Tableau is used for creating interactive and insightful visualizations.

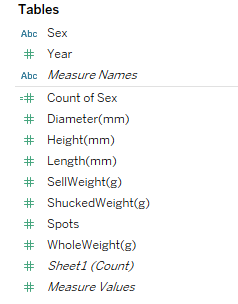


Fig.4. Table Fields

## **Data Visualization in Tableau**

To effectively visualize the King Crab dataset, follow these steps:

## **4.1 Bar Charts**

* Count of crabs by sex and year.
* Average weight (whole, shucked, sell) by sex and year.

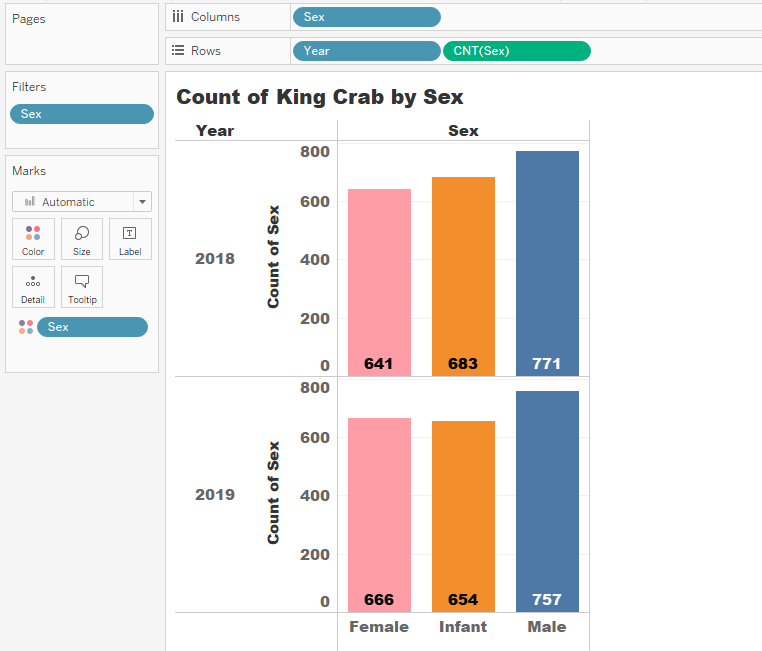


Fig.4.1.1 Count of Crabs by Sex and Year

**Analysis:**

* The male crabs slightly decreased from 2018 to 2019.
* Female crabs increased from 641 in 2018 to 666 in 2019.
* The count of infants decreased from 683 in 2018 to 654 in 2019.

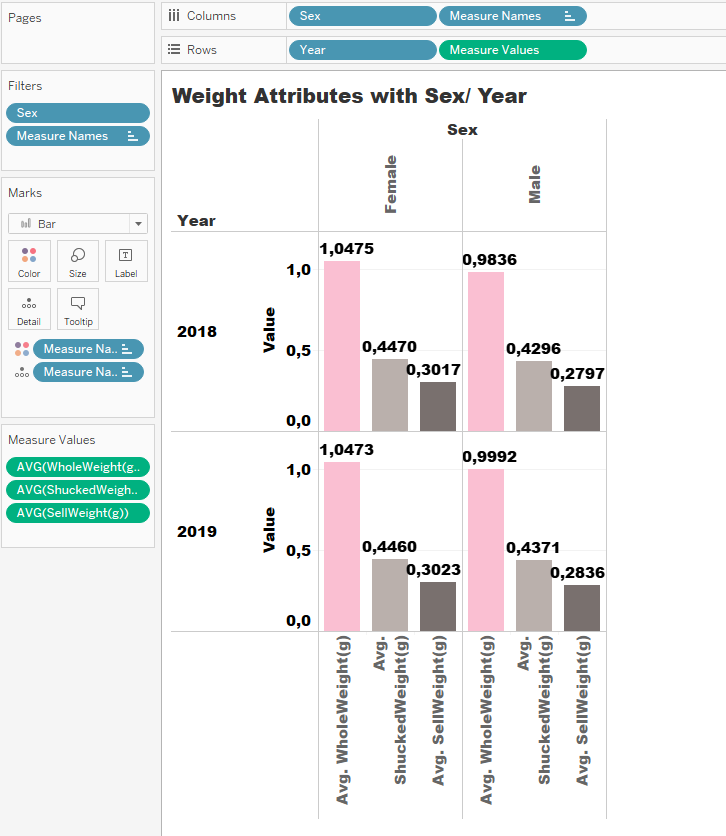


Fig.4.1.2 Average Weight (whole, shucked, sell) by Sex and Year

**Analysis:**

* Female crabs consistently have higher average whole weight, shucked weight, and sell weight compared to male crabs for both years.
* There is a slight increase in the average weights from 2018 to 2019 for both sexes, indicating potential growth in crab sizes.

## **4.2 Line Charts**

* Trends over years for different attributes like length, diameter, height, weights.

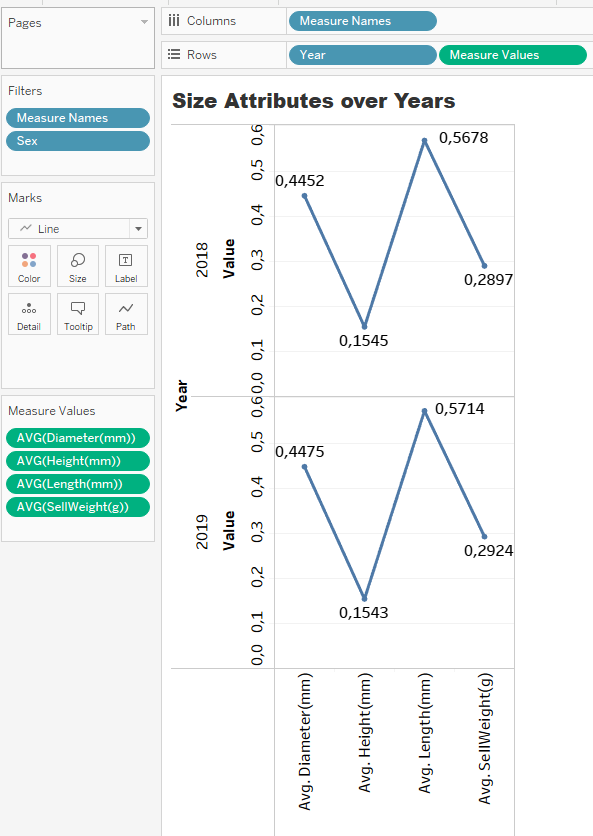


Fig.4.2 Size Attribute over Years

**Analysis:**

**Diameter:**

* Slight increase from 0.4452 in 2018 to 0.4475 in 2019.

**Height:**

* Remained almost constant, with a minor decrease from 0.1545 in 2018 to 0.1543 in 2019.

**Length:**

* Increased from 0.5678 in 2018 to 0.5714 in 2019, indicating a growth trend.

**Sell Weight:**

* Increased from 0.2897 in 2018 to 0.2924 in 2019, suggesting an improvement in the sell weight of crabs.

## **4.3 Scatter Plots**

* Length vs. Weight to understand the correlation.

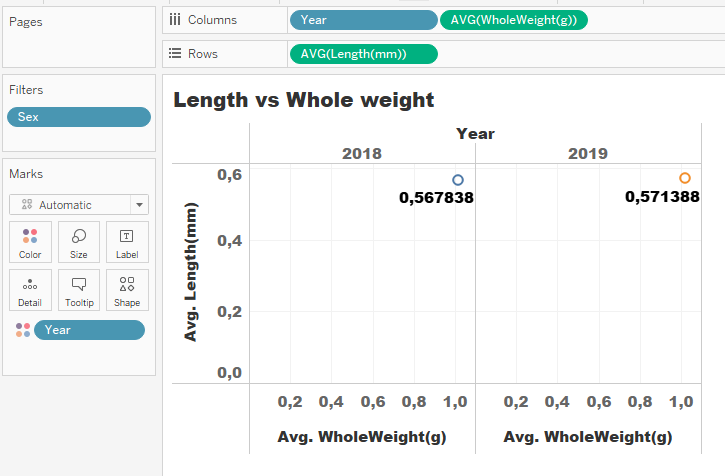


Fig.4.3.1 Lenght vs Whole Weight

**Analysis:**

* **Correlation:**
  + There is a positive correlation between length and whole weight, indicating that larger crabs tend to weigh more.
* **Yearly Comparison:**
  + Comparing the scatter plots for 2018 and 2019 shows that the average length and whole weight increased slightly in 2019 compared to 2018.

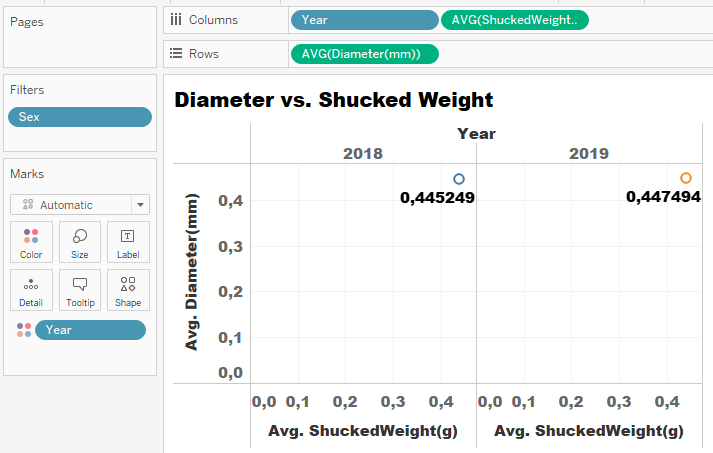


Fig.4.3.2 Diameter vs. Shucked Weight

**Analysis:**

* **Correlation:**
  + There is a positive correlation between diameter and shucked weight, indicating that crabs with larger diameters tend to have more shucked weight.
* **Yearly Comparison:**
  + Comparing the scatter plots for 2018 and 2019 shows a slight increase in both average diameter and average shucked weight in 2019 compared to 2018.

## **4.4 Box and Whisker Plots**

* Distribution of lengths, diameters, and heights by sex and year.

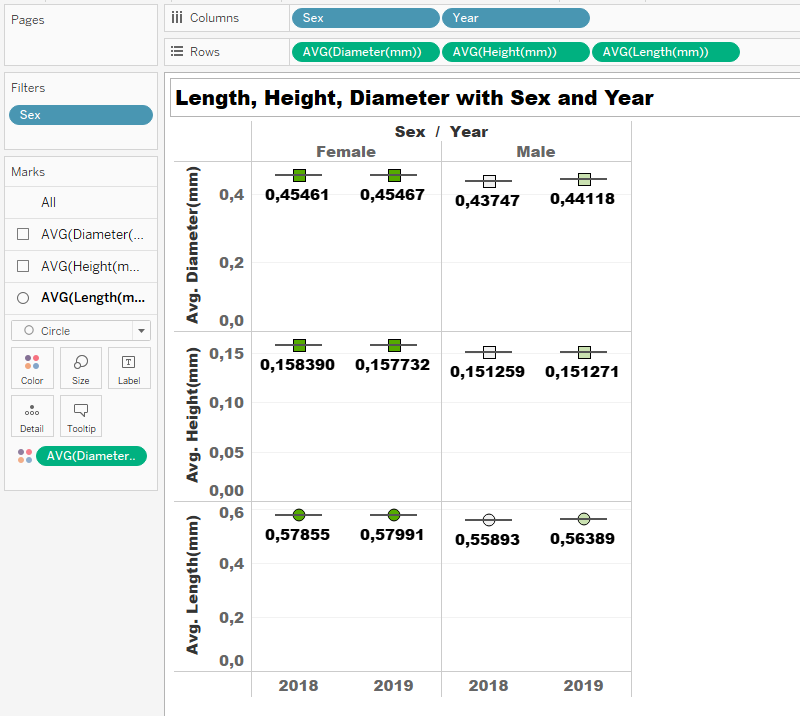


Fig.4.4 Length, Height, Diameter with Sex and Years

**Analysis:**

1. **Diameter Distribution:**
   * Female crabs generally have a larger average diameter compared to male crabs across both years.
   * There is a slight increase in average diameter from 2018 to 2019 for both females and males.
2. **Height Distribution:**
   * Female crabs have a slightly higher average height compared to male crabs in both 2018 and 2019.
   * The height appears to be relatively consistent between the two years for both sexes.
3. **Length Distribution:**
   * Female crabs have a longer average length compared to male crabs in both 2018 and 2019.
   * There is an increase in average length from 2018 to 2019 for both females and males, with females showing a more pronounced increase.

## **4.5 Heat Maps**

* Showing concentration of different weights across various dimensions (e.g., sex and year).
* Heat Map for Whole Weight:

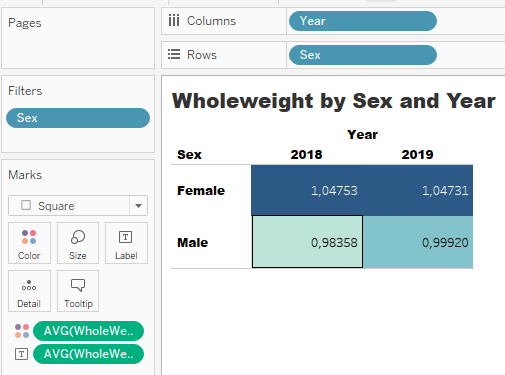


Fig.4.5.1 Heatmap of Whole Weight by Sex and Years

**Analysis:**

1. **Whole Weight Distribution:**
   * Female crabs generally have a higher average whole weight compared to male crabs in both 2018 and 2019.
   * There is a slight decrease in average whole weight for both males and females from 2018 to 2019, with females showing a more noticeable decrease.
   * The darker color indicates higher average weights, which corresponds to the higher values observed for females in both years compared to males.

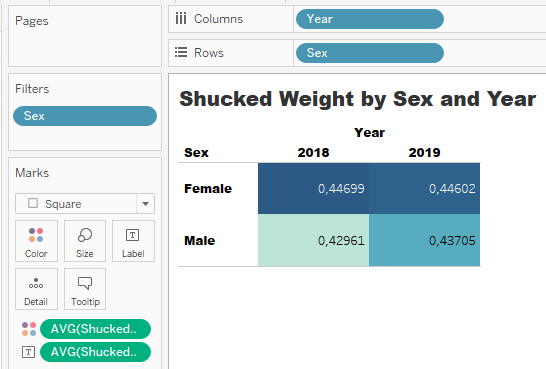


Fig.4.5.2 Heatmap of Shucked Weight by Sex and Years

**Analysis:**

**Shucked Weight Distribution:**

* + Female crabs generally have a higher average shucked weight compared to male crabs in both 2018 and 2019.
  + There is a slight increase in average shucked weight for both males and females from 2018 to 2019, with males showing a more noticeable increase.
  + The darker color indicates higher average weights, which corresponds to the higher values observed for females in both years compared to males.

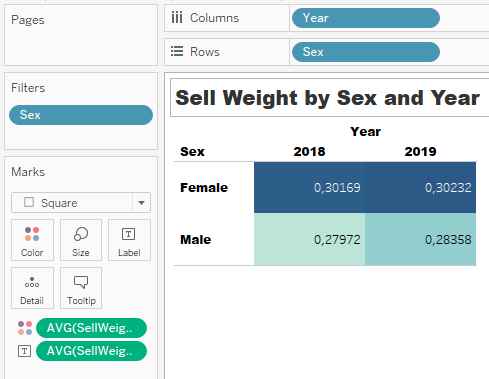


Fig.4.5.3 Heatmap of Sell Weight by Sex and Years

**Analysis:**

**Sell Weight Distribution:**

* + Female crabs generally have a higher average sell weight compared to male crabs in both 2018 and 2019.
  + There is a noticeable increase in average sell weight for both males and females from 2018 to 2019.
  + The darker color on the heat map indicates higher average weights, which aligns with the higher values observed for females compared to males.

## **4.6 Text label**

Sex Ratio Percentage:

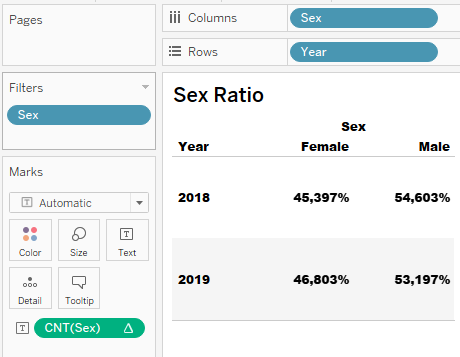


Fig.4.6 Sex Ratio Percentage

**Analysis:**

1. **Yearly Changes:**
   * There is a slight increase in the percentage of female crabs from 2018 to 2019.
   * Conversely, there is a slight decrease in the percentage of male crabs over the same period.

### 5 Storytelling with Data

Narrating information through Tableau provides an effective method for conveying critical insights regarding the King Crab population in Alaska to stakeholders responsible for their welfare and sustainability (Botsis et al., 2020). By visualizing essential indicators, such as growth patterns and demographic trends, Tableau facilitates a deeper understanding of population dynamics and aids in the formulation of targeted management strategies.

To begin, a scatter plot in Tableau illustrates the positive correlation between length (mm) and whole weight (g) of King Crabs. This visualization is pivotal for stakeholders as it predicts weight based on length, essential for population assessments and resource allocation. Understanding this relationship is crucial for assessing population health and size variations over time.

Additionally, a pie chart segmented by sex highlights the distribution of whole weight (g) among male and female King Crabs. This visualization offers insights into population demographics, aiding in the identification of biases or trends in weight distribution. Such insights are instrumental in understanding reproductive patterns and guiding sustainable fishing practices.

These visualizations collectively provide stakeholders, including researchers, fisheries managers, and policymakers, with actionable information presented in a comprehensible format. This clarity supports decision-making processes, whether setting fishing quotas, implementing conservation measures, or revising management policies based on demographic insights.

## **5.1. Implications and Insights**

From the visualizations presented, several critical conclusions emerge that are pivotal for analyzing the dataset and its practical applications:

1. **Weight Distribution by Sex:**
   * There is a noticeable disparity in body weight distribution between male and female King Crabs. This variation may stem from biological factors or sampling effects, underscoring the importance of considering sex-specific aspects in management practices.
2. **Length and Weight Correlation:**
   * The observed positive correlation between length and whole weight signifies a fundamental growth pattern. This relationship is valuable for biomass estimation in marine populations and informs effective management strategies (Juhasz-Dora et al., 2024).
3. **Weight Patterns Across Heights and Sexes:**
   * Variations in weight among different heights and sexes elucidate growth patterns and distribution trends. Identifying denser weight concentrations in specific height classes and sexes informs optimal breeding and harvesting recommendations for maximizing growth and yield (Stevens and Jewett, 2014).
4. **Sex-Based Weight Percentages:**
   * Percentage analysis of weight by sex provides insights into population composition, crucial for assessing demographic structures and guiding intervention measures for conservation and resource utilization.

### 6 Conclusion and Recommendations

**Key Findings:**

* Significant variations in body weight distribution are observed between male and female King Crabs.
* A direct, positive relationship exists between length (mm) and whole weight (g).
* Weight density exhibits non-uniform distribution across different heights and sexes.
* The dispersion of whole weight (g) provides insights into the King Crab population dynamics.

**Recommendations:**

* Develop separate growth models for male and female King Crabs to enhance understanding of growth dynamics.
* Implement targeted breeding and harvesting strategies aligned with height/weight regulations to optimize production.
* Utilize data visualization tools, like Tableau, for informed resource allocation and conservation planning.
* Encourage further research on biological and ecological implications of weight distribution features identified.

In summary, storytelling with data through Tableau transforms complex datasets into accessible and valuable narratives. By bridging the gap between data collection and interpretation, stakeholders gain a nuanced understanding of King Crab population dynamics and the necessary actions for their sustainability and health.

### 7 References

1. Rubin, D. B. (1976). "Inference and missing data." Biometrika, 63(3), 581-592.

2. Kimball, R., & Caserta, J. (2004). "The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data." John Wiley & Sons.

3. van der Aalst, W. M., & Damiani, E. (2015). "Processes meet Big Data: Connecting data science with process science." IEEE Transactions on Services Computing, 8(6), 810-819.

4. Redman, T. C. (1998). "The impact of poor data quality on the typical enterprise." Communications of the ACM, 41(2), 79-82.

5. Tukey, J. W. (1977). "Exploratory Data Analysis." Addison-Wesley.

6. Conover, W. J. (1999). "Practical Nonparametric Statistics." John Wiley & Sons.

7. Chatfield, C. (2003). "The Analysis of Time Series: An Introduction." CRC Press.

8. Rodgers, J. L., & Nicewander, W. A. (1988). "Thirteen ways to look at the correlation coefficient." The American Statistician, 42(1), 59-66.

9. Tufte, E. R. (2001). "The Visual Display of Quantitative Information." Graphics Press.

10. Botsis, T., Hartvigsen, G., Chen, F., Weng, C., & Haug, P. J. (2020). Secondary use of EHR: Data quality issues and informatics opportunities. Summit on Clinical Research Informatics, 2020(1), 234–238. <https://pubmed.ncbi.nlm.nih.gov/33083439/>

11. NOAA Fisheries - Red King Crab. (n.d.). Retrieved from <https://www.fisheries.noaa.gov/species/red-king-crab>

12. Westphal, L. M. (2011). Environmental factors influencing the distribution, abundance, and productivity of the eastern Bering Sea red king crab (Paralithodes camtschaticus). PhD Thesis, University of Washington. https://digital.lib.washington.edu/researchworks/handle/1773/17615

13. Juhasz-Dora, D., Gyetvai, G., Urbar, A., & Szabo, Z. K. (2024). The relationship between size and weight in marine organisms. Marine Biology Research, 20(1), 56-62. https://doi.org/10.1080/17451000.2023.1876312

14. Stevens, B. G., & Jewett, S. C. (2014). Habitat ecology of juvenile red king crab in the southeastern Bering Sea. Deep Sea Research Part II: Topical Studies in Oceanography, 109, 250-259. https://doi.org/10.1016/j.dsr2.2014.02.018