HW #1

1.

a) $\cos(\phi) = \dfrac{x^{(1)} \cdot x^{(2)}}{\underbrace{\|x^{(1)}\|}\underbrace{\|x^{(2)}\|}}$ $\rightarrow = a_1a_1 + a_2a_3 + a_3(-a_2) = a_1^2$

$\rightarrow = \sqrt{a_1^2+a_2^2+a_3^2}$

$= \sqrt{a_1^2+a_2^2+a_3^2}$

$\cos(\phi) = \dfrac{a_1^2}{a_1^2+a_2^2+a_3^2}$

$\boxed{\phi = \cos^{-1}\left(\dfrac{a_1^2}{a_1^2+a_2^2+a_3^2}\right)}$

They are orthogonal when $x^{(1)} \cdot x^{(2)} = 0$, which means $a_1^2 = 0$, so when $a_1 = 0$

b) you can use a non-zero scalar, like c. $(c\theta)\cdot x = c(\theta \cdot x) = 0$. So there are an infinite # of choices to describe the vector, as long as it is not 0.

c)

i) $\dfrac{1\cdot(-1) + (-2)\cdot 1}{\sqrt{1^2+(-2)^2}} = \boxed{\dfrac{-3}{\sqrt{5}}}$

ii) $\dfrac{(1,-2)\cdot(-1,1)+5}{\sqrt{1^2+(-2)^2}} = \boxed{\dfrac{2}{\sqrt{5}}}$

D)

i) you can't uniquely recover the original x just from it's projected point

ii) $x^{(0)} = x - \dfrac{\theta \cdot x + b}{\|\theta\|^2}\theta$, $\{x^{(0)} + \alpha\theta : \alpha \in \mathbb{R}\}$ projects to the same point $x^{(0)}$.

2.

a)

i) $\sum_{i=1}^{m} v_i v_i = v^T v$

ii) $\sum_{i=1}^{m} v_i^2 = \|v\|^2 = v^T v$

iii) $C = vv^T$ $\quad C_{ij} = v_i v_j$

b) $\sum_{i=1}^{n} (\vec{\theta} \cdot \vec{x}^{(i)} - y^{(i)})^2 = \boxed{\|X\theta - y\|^2}$

C) It is not true, the correct traspose rule is $(AB)^T = B^T A^T$

Ex) $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$ $\quad B = \begin{pmatrix} 5 & 6 \\ 7 & 8 \end{pmatrix}$

$AB = \begin{pmatrix} 1\cdot5+2\cdot7 & 1\cdot6+2\cdot8 \\ 3\cdot5 \;\; 4\cdot7 & 3\cdot6+4\cdot8 \end{pmatrix} = \begin{pmatrix} 19 & 22 \\ 43 & 50 \end{pmatrix}$ $\qquad AB^T = \begin{pmatrix} 19 & 43 \\ 22 & 50 \end{pmatrix}$

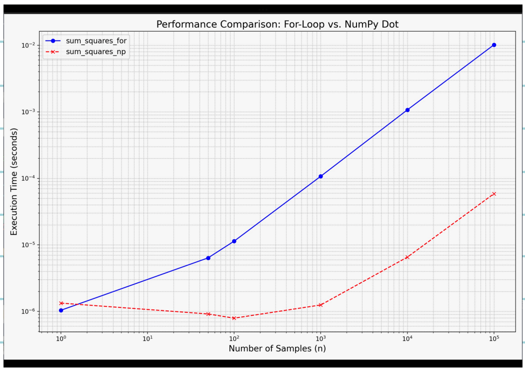
$A^T = \begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix}$ $\quad B^T = \begin{pmatrix} 5 & 7 \\ 6 & 8 \end{pmatrix}$ $\quad A^T B^T = \begin{pmatrix} 1\cdot5+3\cdot6 & 1\cdot7+3\cdot8 \\ 2\cdot5+4\cdot6 & 2\cdot7\cdot4\cdot8 \end{pmatrix} = \begin{pmatrix} 23 & 31 \\ 34 & 46 \end{pmatrix}$

$AB^T = \begin{pmatrix} 19 & 43 \\ 22 & 50 \end{pmatrix} \neq \begin{pmatrix} 22 & 31 \\ 34 & 46 \end{pmatrix} = A^T B^T$

D) It is not possible because $\text{rank}(AB) < \min(\text{rank}(A), \text{rank}(B)) \leq p < n$, so $C = AB$ cannot be full rank, so rank C cannot reach n.

Q3)

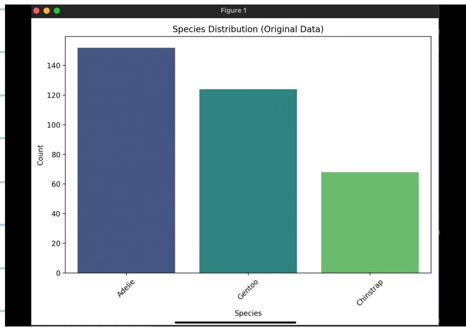
f)

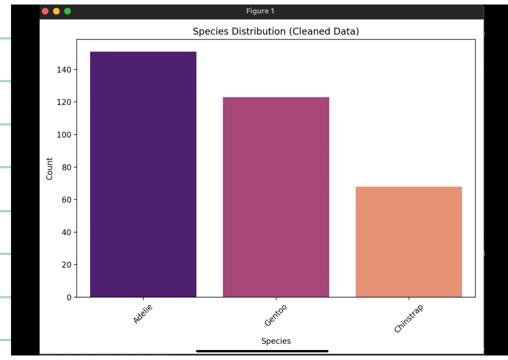



```python
def plot_timings(time_df):
    plt.figure(figsize=(12, 8))
    plt.plot(time_df['n'], time_df['ssfor'], label='sum_squares_for', marker='o', linestyle='-', color='blue')
    plt.plot(time_df['n'], time_df['ssnp'], label='sum_squares_np', marker='x', linestyle='--', color='red')
    plt.xlabel('Number of Samples (n)', fontsize=14)
    plt.ylabel('Execution Time (seconds)', fontsize=14)
    plt.title('Performance Comparison: For-Loop vs. NumPy Dot', fontsize=16)
    plt.legend(fontsize=12)
    plt.grid(True, which="both", ls="--", linewidth=0.5)
    plt.xscale('log')
    plt.yscale('log')
    plt.tight_layout()
    plt.show()
```

Q5)

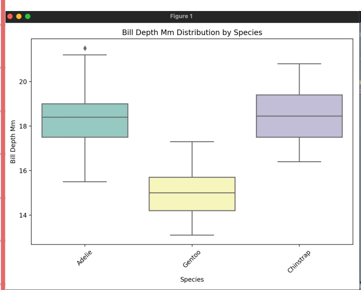
a)

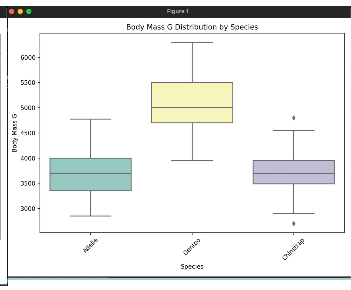


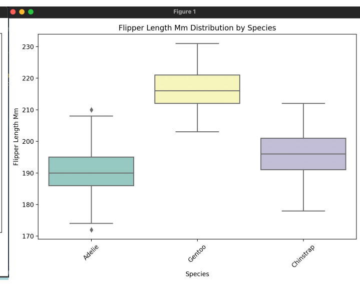


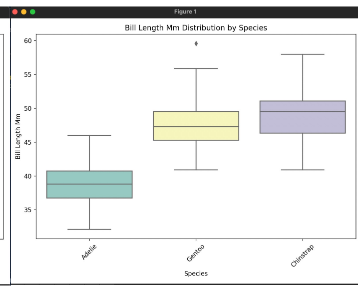
Dropping the data with incomplete measurements does not change the species distribution because when it removes rows with missing numeric data, it does not directly impact the species column. So the species count should remain proportional.

b)

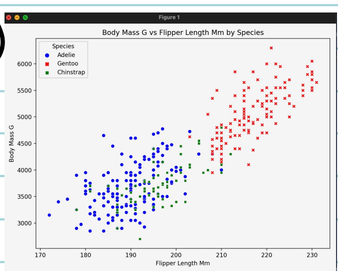


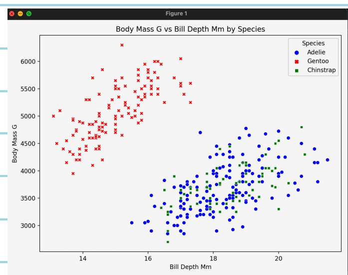


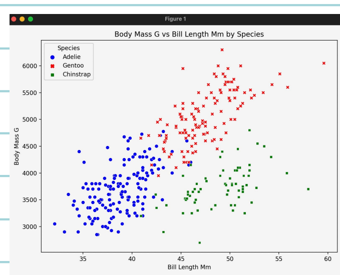


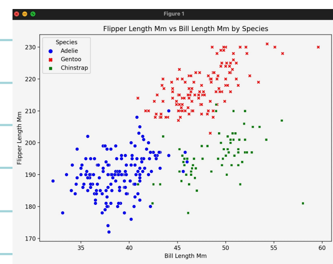


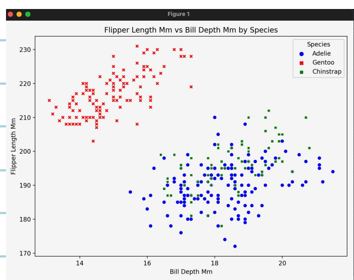
C)

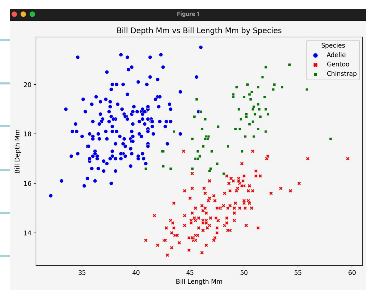












D) one rule could be that for the body mass threshold, if a penguin has a body mass that is more than 4000 grams, it would most likely be classfied as a Gentoo. Another rule could be for flipper length measurement. If the length is greater than 200 mm, it can most likely be classified as a Gentoo. Another rule can be for Bill length and depth combination. If the bill length is more than 45 mm, and the bill depth is greater than 18 mm, it can be most likely classified as a chinstrap. The last rule could be for default classification. For any penguins that have a short bill and flippers, they are most likely Adelles.