- **Q1**. How many observations are in this dataset?

  569 observations

- **Q2**. How many of the observations have a malignant diagnosis?

  212 malignant diagnoses

- **Q3**. How many variables/features in the data are suffixed with _mean?

  10 variables

- **Q4**. From your results, what proportion of the original variance is captured by the first principal components (PC1)?

  98.2%

- **Q5**. How many principal components (PCs) are required to describe at least 70% of the original variance in the data?

  1

- **Q6**. How many principal components (PCs) are required to describe at least 90% of the original variance in the data?

  1

- **Q7.** What stands out to you about this plot? Is it easy or difficult to understand? Why

  This biplot is unreadable and needs to be cleaned up in order to understand. All the numbers and text is on top of each other.

- **Q8.** Generate a similar plot for principal components 1 and 3.
  What do you notice about these plots?

  The diagnosis B tends towards the right of the plot with diagnosis M towards the left.

- **Q9.** For the first principal component, what is the component of the loading vector (i.e. `wisc.pr$rotation[,1]`) for the feature `concave.points_mean`?

- **Q10.** What is the minimum number of principal components required to explain 80% of the variance of the data?

  1 principal component

- **Q11.** Using the `plot()` and `abline()` functions, what is the height at which the clustering model has 4 clusters?

  19

- **Q12.** Can you find a better cluster vs diagnoses match by cutting into a different number of clusters between 2 and 10?

  Yeah, you can probably find a better cluster match with 2 clusters with one for malignant cells and one fore benign

- **Q13.** Which method gives your favorite results for the same data.dist dataset? Explain your reasoning.

  I like the “complete” method because it gives you everything and from there you can interpret the clustering procedure

- Q18. Which of these new patients should we prioritize for follow up based on your results?

  We should prioritize patient 1