---

title: “Mini_project_SASSE” author: “Noah Sasse” date: “2/16/2022” output: pdf_document: default html_document: default

wisc.df <- WisconsinCancer wisc.data <- wisc.df[,-1] diagnosis <- wisc.data[,1] colMeans(wisc.data[,-1]) apply(wisc.data,2,sd) wisc.pr <- prcomp(wisc.data[,-1]) summary(wisc.pr)’

biplot(wisc.pr) plot(wisc.pr$x, col = diagnosis, xlab = “PC1”, ylab = “PC2”)

- **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.