```
% Options for packages loaded elsewhere
\PassOptionsToPackage{unicode}{hyperref}
\PassOptionsToPackage{hyphens}{url}
%
\documentclass[
]{article}
\usepackage{lmodern}
\usepackage{amsmath}
\usepackage{ifxetex,ifluatex}
\ifnum 0\ifxetex 1\fi\ifluatex 1\fi=0 % if pdftex
  \usepackage[T1]{fontenc}
  \usepackage[utf8]{inputenc}
  \usepackage{textcomp} % provide euro and other symbols
  \usepackage{amssymb}
\else % if luatex or xetex
  \usepackage{unicode-math}
  \defaultfontfeatures{Scale=MatchLowercase}
  \defaultfontfeatures[\rmfamily]{Ligatures=TeX,Scale=1}
\fi
% Use upquote if available, for straight quotes in verbatim environments
\IfFileExists{upquote.sty}{\usepackage{upquote}}{}
\IfFileExists{microtype.sty}{% use microtype if available
  \usepackage[]{microtype}
  \UseMicrotypeSet[protrusion]{basicmath} % disable protrusion for tt fonts
}{}
\makeatletter
\@ifundefined{KOMAClassName}{% if non-KOMA class
  \IfFileExists{parskip.sty}{%
    \usepackage{parskip}
  }{% else
    \setlength{\parindent}{0pt}
    \setlength{\parskip}{6pt plus 2pt minus 1pt}}
}{% if KOMA class
  \KOMAoptions{parskip=half}}
\makeatother
\usepackage{xcolor}
\IfFileExists{xurl.sty}{\usepackage{xurl}}{} % add URL line breaks if available
\IfFileExists{bookmark.sty}{\usepackage{bookmark}}{\usepackage{hyperref}}
\hypersetup{
  pdftitle={Lab 2},
  pdfauthor={Hanlin Wang},
  hidelinks,
  pdfcreator={LaTeX via pandoc}}
\urlstyle{same} % disable monospaced font for URLs
\usepackage[margin=1in]{geometry}
\usepackage{color}
\usepackage{fancyvrb}
\newcommand{\VerbBar}{|}
\newcommand{\VERB}{\Verb[commandchars=\\\{\}]}
\DefineVerbatimEnvironment{Highlighting}{Verbatim}{commandchars=\\\{\}}
% Add ',fontsize=\small' for more characters per line
\usepackage{framed}
\definecolor{shadecolor}{RGB}{248,248,248}
\newenvironment{Shaded}{\begin{snugshade}}{\end{snugshade}}
\newcommand{\AlertTok}[1]{\textcolor[rgb]{0.94,0.16,0.16}{#1}}
\newcommand{\AnnotationTok}[1]{\textcolor[rgb]{0.56,0.35,0.01}{\textbf{\textit{#1}}}}
\newcommand{\AttributeTok}[1]{\textcolor[rgb]{0.77,0.63,0.00}{#1}}
```

```
\newcommand{\BaseNTok}[1]{\textcolor[rgb]{0.00,0.00,0.81}{#1}}
\newcommand{\BuiltInTok}[1]{#1}
\newcommand{\CharTok}[1]{\textcolor[rgb]{0.31,0.60,0.02}{#1}}
\newcommand{\CommentTok}[1]{\textcolor[rgb]{0.56,0.35,0.01}{\textit{#1}}}
\newcommand{\CommentVarTok}[1]{\textcolor[rgb]{0.56,0.35,0.01}{\textbf{\textit{#1}}}}
\newcommand{\ConstantTok}[1]{\textcolor[rgb]{0.00,0.00,0.00}{#1}}
\newcommand{\ControlFlowTok}[1]{\textcolor[rgb]{0.13,0.29,0.53}{\textbf{#1}}}
\newcommand{\DataTypeTok}[1]{\textcolor[rgb]{0.13,0.29,0.53}{#1}}
\newcommand{\DecValTok}[1]{\textcolor[rgb]{0.00,0.00,0.81}{#1}}
\newcommand{\DocumentationTok}[1]{\textcolor[rgb]{0.56,0.35,0.01}{\textbf{\textit{#1}}}}
\newcommand{\ErrorTok}[1]{\textcolor[rgb]{0.64,0.00,0.00}{\textbf{#1}}}
\newcommand{\ExtensionTok}[1]{#1}
\newcommand{\FloatTok}[1]{\textcolor[rgb]{0.00,0.00,0.81}{#1}}
\newcommand{\FunctionTok}[1]{\textcolor[rgb]{0.00,0.00,0.00}{#1}}
\newcommand{\ImportTok}[1]{#1}
\newcommand{\InformationTok}[1]{\textcolor[rgb]{0.56,0.35,0.01}{\textbf{\textit{#1}}}}
\newcommand{\KeywordTok}[1]{\textcolor[rgb]{0.13,0.29,0.53}{\textbf{#1}}}
\newcommand{\NormalTok}[1]{#1}
\newcommand{\OperatorTok}[1]{\textcolor[rgb]{0.81,0.36,0.00}{\textbf{#1}}}
\newcommand{\OtherTok}[1]{\textcolor[rgb]{0.56,0.35,0.01}{#1}}
\newcommand{\PreprocessorTok}[1]{\textcolor[rgb]{0.56,0.35,0.01}{\textit{#1}}}
\newcommand{\RegionMarkerTok}[1]{#1}
\newcommand{\SpecialCharTok}[1]{\textcolor[rgb]{0.00,0.00,0.00}{#1}}
\newcommand{\SpecialStringTok}[1]{\textcolor[rgb]{0.31,0.60,0.02}{#1}}
\newcommand{\StringTok}[1]{\textcolor[rgb]{0.31,0.60,0.02}{#1}}
\newcommand{\VariableTok}[1]{\textcolor[rgb]{0.00,0.00,0.00}{#1}}
\newcommand{\VerbatimStringTok}[1]{\textcolor[rgb]{0.31,0.60,0.02}{#1}}
\newcommand{\WarningTok}[1]{\textcolor[rgb]{0.56,0.35,0.01}{\textbf{\textit{#1}}}}
\usepackage{longtable,booktabs}
\usepackage{calc} % for calculating minipage widths
% Correct order of tables after \paragraph or \subparagraph
\usepackage{etoolbox}
\makeatletter
\patchcmd\longtable{\par}{\if@noskipsec\mbox{}\fi\par}{}{}
\makeatother
% Allow footnotes in longtable head/foot
\IfFileExists{footnotehyper.sty}{\usepackage{footnotehyper}}{\usepackage{footnote}}
\makesavenoteenv{longtable}
\usepackage{graphicx}
\makeatletter
\def\maxwidth{\ifdim\Gin@nat@width>\linewidth\linewidth\else\Gin@nat@width\fi}
\def\maxheight{\ifdim\Gin@nat@height>\textheight\textheight\else\Gin@nat@height\fi}
\makeatother
% Scale images if necessary, so that they will not overflow the page
% margins by default, and it is still possible to overwrite the defaults
% using explicit options in \includegraphics[width, height, ...]{}
\setkeys{Gin}{width=\maxwidth,height=\maxheight,keepaspectratio}
% Set default figure placement to htbp
\makeatletter
\def\fps@figure{htbp}
\makeatother
\setlength{\emergencystretch}{3em} % prevent overfull lines
\providecommand{\tightlist}{%
  \setlength{\itemsep}{0pt}\setlength{\parskip}{0pt}}
\setcounter{secnumdepth}{-\maxdimen} % remove section numbering
\ifluatex
  \usepackage{selnolig}  % disable illegal ligatures
```

```
\fi

\title{Lab 2}
\author{Hanlin Wang}
\date{11:59PM February 25, 2021}

\begin{document}
\maketitle

\# More Basic R Skills

\begin{itemize}
\tightlist
\item
   Create a function \texttt{my\_reverse} which takes as required input a
   vector and returns the vector in reverse where the first entry is the
   last entry, etc. No function calls are allowed inside your function
   otherwise that would defeat the purpose of the exercise! (Yes, there
   is a base R function that does this called \texttt{rev}). Use
   \texttt{head} on \texttt{v} and \texttt{tail} on
   \texttt{my\_reverse(v)} to verify it works.
\end{itemize}

\begin{Shaded}
\begin{Highlighting}[]
\NormalTok{my\_reverse }\OtherTok{=} \ControlFlowTok{function}\NormalTok{(v)\{}
\NormalTok{    v\_rev }\OtherTok{=}
\FunctionTok{rep}\NormalTok{(}\ConstantTok{NA}\NormalTok{, }\AttributeTok{times =}
\FunctionTok{length}\NormalTok{(v))}
   \ControlFlowTok{for}\NormalTok{ (i }\ControlFlowTok{in}
\DecValTok{1}\SpecialCharTok{:}\FunctionTok{length}\NormalTok{(v)) \{}
\NormalTok{        v\_rev[}\FunctionTok{length}\NormalTok{(v) }\SpecialCharTok{-}\NormalTok{ i }\S
pecialCharTok{+} \DecValTok{1}\NormalTok{] }\OtherTok{=}\NormalTok{ v[i]}
\NormalTok{    \}}

\NormalTok{    v\_rev}
\NormalTok{\}}

\NormalTok{v }\OtherTok{=} \DecValTok{1} \SpecialCharTok{:} \DecValTok{10}

\FunctionTok{my\_reverse}\NormalTok{(v)}
\end{Highlighting}
\end{Shaded}

\begin{verbatim}
##    [1] 10    9    8    7    6    5    4    3    2    1
\end{verbatim}

\begin{itemize}
\tightlist
\item
   Create a function \texttt{flip\_matrix} which takes as required input
   a matrix, an argument \texttt{dim\_to\_rev} that returns the matrix
   with the rows in reverse order or the columns in reverse order
   depending on the \texttt{dim\_to\_rev} argument. Let the default be
   the dimension of the matrix that is greater.
\end{itemize}
```

```
\begin{Shaded}
\begin{Highlighting}[]
\NormalTok{flip\_matrix }\OtherTok{=}
\ControlFlowTok{function}\NormalTok{(}\AttributeTok{dim\_to\_rev =}
\ConstantTok{NULL}\NormalTok{) \{}
   \ControlFlowTok{if}\NormalTok{ (}\FunctionTok{is.null}\NormalTok{(dim\_to\_rev))\{}
\NormalTok{       dim\_to\_rev }\OtherTok{=}
\FunctionTok{ifelse}\NormalTok{ (}\FunctionTok{nrow}\NormalTok{(x) }\SpecialCharTok{\textgreater{}}
=}
\FunctionTok{ncol}\NormalTok{(x), }\StringTok{"rows"}\NormalTok{, }\StringTok{"cols"}\NormalTok{)}
\NormalTok{    \}}
   \ControlFlowTok{if}\NormalTok{ (dim\_to\_rev }\SpecialCharTok{==}
\StringTok{"cols"}\NormalTok{)\{}
\NormalTok{       x[}\FunctionTok{my\_reverse}\NormalTok{(}\DecValTok{1}\SpecialCharTok{:}\Functi
onTok{norw}\NormalTok{(x)) , ]    }
\NormalTok{    \}}
   \ControlFlowTok{else} \ControlFlowTok{if}\NormalTok{ (dim\_to\_rev }\SpecialCharTok{==}
\StringTok{"cols"}\NormalTok{)\{}
\NormalTok{       x[ , }\FunctionTok{my\_reverse}\NormalTok{(}\DecValTok{1}\SpecialCharTok{:}\Fun
ctionTok{ncol}\NormalTok{(x))]    }
\NormalTok{    \}}
   \ControlFlowTok{else}\NormalTok{\{}
      \FunctionTok{stop}\NormalTok{(}\StringTok{"illegal argument"}\NormalTok{)}
\NormalTok{    \}}
\NormalTok{\}}
\NormalTok{x }\OtherTok{=}
\FunctionTok{matrix}\NormalTok{(}\FunctionTok{rnorm}\NormalTok{(}\DecValTok{100}\NormalTok{),
 }\AttributeTok{nrow =} \DecValTok{25}\NormalTok{)}
\NormalTok{x}
\end{Highlighting}
\end{Shaded}

\begin{verbatim}
##                [,1]       [,2]       [,3]       [,4]
##  [1,] -0.03912137 -1.03390093  0.04062372 -0.3802549
##  [2,]  1.08879392  0.23496225 -0.34171902  0.6296714
##  [3,] -0.43374453 -1.28996467  0.90879783 -0.5608175
##  [4,]  0.96535930 -0.88009139  1.07441464  0.1881048
##  [5,] -0.63591995 -0.86237847 -0.49130655  0.4643298
##  [6,]  0.79892535  1.23158237  0.76048590 -0.1091033
##  [7,] -2.29389104  1.59263195 -1.71866668  0.3204528
##  [8,] -0.23083127  0.78361506  0.87222600  0.9531006
##  [9,] -0.72045438 -0.04686282 -0.13502825  2.9603479
## [10,]  0.35988972 -0.79156034  0.03062629  1.2585087
## [11,] -0.07187930 -0.49464127  0.10363010  0.4419800
## [12,]  0.48195469 -0.23523080 -0.95614400 -0.7283575
## [13,] -1.77657856  0.32235742  0.31139587 -1.1492272
## [14,]  0.42517659 -1.14587167 -0.80002449 -1.8983196
## [15,] -0.65370067 -0.38682544 -0.22417664 -1.6787377
## [16,]  0.63802737  1.21564819 -0.95155613  0.5234012
## [17,]  0.99272197  1.09568171  1.24267060  0.0512680
## [18,]  0.19584755 -0.24875014  1.50864507 -0.4211908
## [19,]  0.19552534 -0.83594125  0.19789606 -0.3522275
## [20,]  0.02237697 -1.24607007 -0.79251729  0.6530997
## [21,]  0.12958023 -0.80235820 -0.07140539 -2.0564505
## [22,] -0.22072310 -0.56599869 -0.67736418 -0.6491193
```

```
## [23,] -0.46256612 -0.59171584 -0.25705350 -0.4596104
## [24,] -0.20052778  0.79065883 -0.72513645  0.5332064
## [25,]  0.05976847 -1.15971860 -0.59178045 -1.0176038
\end{verbatim}

\begin{Shaded}
\begin{Highlighting}[]
\NormalTok{flip\_matrix}
\end{Highlighting}
\end{Shaded}

\begin{verbatim}
## function(dim_to_rev = NULL) {
##     if (is.null(dim_to_rev)){
##         dim_to_rev = ifelse (nrow(x) >= ncol(x), "rows", "cols")
##     }
##     if (dim_to_rev == "cols"){
##         x[my_reverse(1:norw(x)) , ]
##     }
##     else if (dim_to_rev == "cols"){
##         x[ , my_reverse(1:ncol(x))]
##     }
##     else{
##         stop("illegal argument")
##     }
## }
\end{verbatim}

\begin{itemize}
\tightlist
\item
  Create a list named \texttt{my\_list} with keys ``A'', ``B'', \ldots{}
  where the entries are arrays of size 1, 2 x 2, 3 x 3 x 3, etc. Fill
  the array with the numbers 1, 2, 3, etc. Make 8 entries according to
  this sequence.
\end{itemize}

\begin{Shaded}
\begin{Highlighting}[]
\NormalTok{arrays }\OtherTok{=} \FunctionTok{list}\NormalTok{()}

\NormalTok{LETTERS}
\end{Highlighting}
\end{Shaded}

\begin{verbatim}
##   [1] "A" "B" "C" "D" "E" "F" "G" "H" "I" "J" "K" "L" "M" "N" "O" "P" "Q" "R" "S"
## [20] "T" "U" "V" "W" "X" "Y" "Z"
\end{verbatim}

\begin{Shaded}
\begin{Highlighting}[]
\NormalTok{arrays[[}\StringTok{"A"}\NormalTok{]] }\OtherTok{=}
\FunctionTok{array}\NormalTok{(}\AttributeTok{data =}
\DecValTok{1}\SpecialCharTok{:}\DecValTok{4}\NormalTok{, }\AttributeTok{dim =}
\FunctionTok{c}\NormalTok{(}\DecValTok{2}\NormalTok{, }\DecValTok{2}\NormalTok{)) }\CommentT
ok{\# Code inside the for loop }
```

```
\NormalTok{arrays}
\end{Highlighting}
\end{Shaded}

\begin{verbatim}
## $A
##      [,1] [,2]
## [1,]    1    3
## [2,]    2    4
\end{verbatim}

\begin{Shaded}
\begin{Highlighting}[]
\FunctionTok{array}\NormalTok{(}\AttributeTok{data =}
\DecValTok{1}\SpecialCharTok{:}\DecValTok{27}\NormalTok{, }\AttributeTok{dim =}
\FunctionTok{c}\NormalTok{(}\DecValTok{3}\NormalTok{, }\DecValTok{3}\NormalTok{, }\DecValTok{3}
\NormalTok{))}
\end{Highlighting}
\end{Shaded}

\begin{verbatim}
## , , 1
##
##      [,1] [,2] [,3]
## [1,]    1    4    7
## [2,]    2    5    8
## [3,]    3    6    9
##
## , , 2
##
##      [,1] [,2] [,3]
## [1,]   10   13   16
## [2,]   11   14   17
## [3,]   12   15   18
##
## , , 3
##
##      [,1] [,2] [,3]
## [1,]   19   22   25
## [2,]   20   23   26
## [3,]   21   24   27
\end{verbatim}

\begin{Shaded}
\begin{Highlighting}[]
\CommentTok{\#for (variable in vector) \{}


\CommentTok{\#\}          }
\end{Highlighting}
\end{Shaded}

Run the following code:

\begin{Shaded}
\begin{Highlighting}[]
\CommentTok{\#lapply(my\_list, object.size)}
```

\end{Highlighting}
\end{Shaded}

Use \texttt{?object.size} to read about what these functions do. Then explain the output you see above. For the later arrays, does it make sense given the dimensions of the arrays?

\#TO-DO

Now cleanup the namespace by deleting all stored objects and functions:

\begin{Shaded}
\begin{Highlighting}[]
\CommentTok{\#TO{-}DO}
\end{Highlighting}
\end{Shaded}

\hypertarget{a-little-about-strings}{%
\subsection{A little about strings}\label{a-little-about-strings}}

\begin{itemize}
\tightlist
\item
  Use the \texttt{strsplit} function and \texttt{sample} to put the sentences in the string \texttt{lorem} below in random order. You will also need to manipulate the output of \texttt{strsplit} which is a list. You may need to learn basic concepts of regular expressions.
\end{itemize}

\begin{Shaded}
\begin{Highlighting}[]
\NormalTok{lorem }\OtherTok{=} \StringTok{"Lorem ipsum dolor sit amet, consectetur adipiscing elit. Morbi posuere varius volutpat. Morbi faucibus ligula id massa ultricies viverra. Donec vehicula sagittis nisi non semper. Donec at tempor erat. Integer dapibus mi lectus, eu posuere arcu ultricies in. Cras suscipit id nibh lacinia elementum. Curabitur est augue, congue eget quam in, scelerisque semper magna. Aenean nulla ante, iaculis sed vehicula ac, finibus vel arcu. Mauris at sodales augue. "}
\end{Highlighting}
\end{Shaded}

You have a set of names divided by gender (M / F) and generation (Boomer / GenX / Millenial):

\begin{itemize}
\tightlist
\item
  M / Boomer ``Theodore, Bernard, Gene, Herbert, Ray, Tom, Lee, Alfred, Leroy, Eddie''
\item
  M / GenX ``Marc, Jamie, Greg, Darryl, Tim, Dean, Jon, Chris, Troy, Jeff''
\item
  M / Millennial ``Zachary, Dylan, Christian, Wesley, Seth, Austin, Gabriel, Evan, Casey, Luis''
\item
  F / Boomer ``Gloria, Joan, Dorothy, Shirley, Betty, Dianne, Kay, Marjorie, Lorraine, Mildred''
\item

```
   F / GenX ``Tracy, Dawn, Tina, Tammy, Melinda, Tamara, Tracey, Colleen,
   Sherri, Heidi''
\item
   F / Millennial ``Samantha, Alexis, Brittany, Lauren, Taylor, Bethany,
   Latoya, Candice, Brittney, Cheyenne''
\end{itemize}

Create a list-within-a-list that will intelligently store this data.

\begin{Shaded}
\begin{Highlighting}[]
\CommentTok{\#HINT: }
\CommentTok{\#strsplit("Theodore, Bernard, Gene, Herbert, Ray, Tom, Lee, Alfred, Leroy, Eddie",
split = ", ")[[1]]}
\CommentTok{\#TO{-}DO}
\end{Highlighting}
\end{Shaded}

\hypertarget{dataframe-creation}{%
\subsection{Dataframe creation}\label{dataframe-creation}}

Imagine you are running an experiment with many manipulations. You have
14 levels in the variable ``treatment'' with levels a, b, c, etc. For
each of those manipulations you have 3 submanipulations in a variable
named ``variation'' with levels A, B, C. Then you have ``gender'' with
levels M / F. Then you have ``generation'' with levels Boomer, GenX,
Millenial. Then you will have 6 runs per each of these groups. In each
set of 6 you will need to select a name without duplication from the
appropriate set of names (from the last question). Create a data frame
with columns treatment, variation, gender, generation, name and y that
will store all the unique unit information in this experiment. Leave y
empty because it will be measured as the experiment is executed.

\begin{Shaded}
\begin{Highlighting}[]
\NormalTok{n }\OtherTok{=} \DecValTok{14} \SpecialCharTok{*} \DecValTok{3} \SpecialCharTok{*}
\DecValTok{2} \SpecialCharTok{*} \DecValTok{3} \SpecialCharTok{*} \DecValTok{10}
\CommentTok{\#X = data.frame(treatment = rep(NA,n), }
\CommentTok{\#                            ...}
\CommentTok{\#TO{-}DO}
\end{Highlighting}
\end{Shaded}

\hypertarget{packages}{%
\subsection{Packages}\label{packages}}

Install the package \texttt{pacman} using regular base R.

\begin{Shaded}
\begin{Highlighting}[]
\CommentTok{\# install.packages("pacman")}
\end{Highlighting}
\end{Shaded}

First, install the package \texttt{testthat} (a widely accepted testing
suite for R) from \url{https://github.com/r-lib/testthat} using
\texttt{pacman}. If you are using Windows, this will be a long install,
```

but you have to go through it for some of the stuff we are doing in class. LINUX (or MAC) is preferred for coding. If you can't get it to work, install this package from CRAN (still using \texttt{pacman}), but this is not recommended long term.

```
\begin{Shaded}
\begin{Highlighting}[]
\NormalTok{pacman}\SpecialCharTok{::}\FunctionTok{p\_load}\NormalTok{(testthat)}
\end{Highlighting}
\end{Shaded}
```

```
\begin{itemize}
\tightlist
\item
  Create vector \texttt{v} consisting of all numbers from -100 to 100
  and test using the second line of code su
\end{itemize}
```

```
\begin{Shaded}
\begin{Highlighting}[]
\CommentTok{\# v= seq({-}100, 100)}
\CommentTok{\# expect\_equal(v, {-}100 : 101)}
\end{Highlighting}
\end{Shaded}
```

If there are any errors, the \texttt{expect_equal} function will tell you about them. If there are no errors, then it will be silent.

Test the \texttt{my_reverse} function from lab2 using the following code:

```
\begin{Shaded}
\begin{Highlighting}[]
\CommentTok{\#v = 1:100}
\CommentTok{\#my\_reverse(v)}
\CommentTok{\#rev(v)}

\FunctionTok{expect\_equal}\NormalTok{(}\FunctionTok{my\_reverse}\NormalTok{(v), }\FunctionTok
{rev}\NormalTok{(v))}
\FunctionTok{expect\_equal}\NormalTok{(}\FunctionTok{my\_reverse}\NormalTok{(}\FunctionTok{c}
\NormalTok{(}\StringTok{"A"}\NormalTok{, }\StringTok{"B"}\NormalTok{, }\StringTok{"C"}\NormalTok
{)), }\FunctionTok{c}\NormalTok{(}\StringTok{"C"}\NormalTok{, }\StringTok{"B"}\NormalTok{, }\String
Tok{"A"}\NormalTok{))}
\end{Highlighting}
\end{Shaded}
```

```
\hypertarget{multinomial-classification-using-knn}{%
\subsection{Multinomial Classification using
KNN}\label{multinomial-classification-using-knn}}
```

Write a $k=1$ nearest neighbor algorithm using the Euclidean distance function. This is standard ``Roxygen'' format for documentation. Hopefully, we will get to packages at some point and we will go over this again. It is your job also to fill in this documentation.

```
\begin{Shaded}
\begin{Highlighting}[]
```

```
\CommentTok{\#\textquotesingle{} TO{-}DO: Provide a name for this function                }
\CommentTok{\#\textquotesingle{} one nearest neighbour classifer                          }

\CommentTok{\#\textquotesingle{} TO{-}DO: Explain what this function does in a few sentences}
\CommentTok{\#\textquotesingle{} classify an observation based on the label of the closet
observation in the set of training observation}

\CommentTok{\#\textquotesingle{} @param xinput          TO{-}DO: explain this argument          A
matrix of featrues for training data observation}
\CommentTok{\#\textquotesingle{} @param y\_binary       TO{-}DO: explain this argument
The vector of training data}
\CommentTok{\#\textquotesingle{} @param xtest          TO{-}DO: explain this argument          A
test observation as a row vector}
\CommentTok{\#\textquotesingle{} @return               TO{-}DO: explain this argument
The predicted label for the test observation}
\NormalTok{nn\_algorithm\_predict }\OtherTok{=} \ControlFlowTok{function}\NormalTok{(xinput,
y\_binary, xtest)\{}
\NormalTok{    distances }\OtherTok{=}
\FunctionTok{array}\NormalTok{ (}\ConstantTok{NA}\NormalTok{, n)}
\NormalTok{    n }\OtherTok{=} \FunctionTok{nrow}\NormalTok{(xinput)}
   \ControlFlowTok{for}\NormalTok{(i }\ControlFlowTok{in}
\DecValTok{1}\SpecialCharTok{:}\NormalTok{n)\{}
\NormalTok{       distances[i]}\OtherTok{=}\NormalTok{(}\FunctionTok{sum}\NormalTok{((xinput[i, ] }
\SpecialCharTok{{-}}\NormalTok{ xtest)}\SpecialCharTok{\^{}}\DecValTok{2}\NormalTok{))}
\NormalTok{    \}    }

\NormalTok{    y\_binary[}\FunctionTok{which.min}\NormalTok{(distances)]}

\NormalTok{\}}
\end{Highlighting}
\end{Shaded}
```

Write a few tests to ensure it actually works:

```
\begin{Shaded}
\begin{Highlighting}[]
\CommentTok{\# nn\_algorithm\_predict(iris)}
\CommentTok{\#TO{-}DO}
\end{Highlighting}
\end{Shaded}
```

We now add an argument \texttt{d} representing any legal distance function to the \texttt{nn_algorithm_predict} function. Update the implementation so it performs NN using that distance function. Set the default function to be the Euclidean distance in the original function. Also, alter the documentation in the appropriate places.

```
\begin{Shaded}
\begin{Highlighting}[]
\CommentTok{\#\textquotesingle{} TO{-}DO: Provide a name for this function                }
\CommentTok{\#\textquotesingle{} one nearest neighbour classifer                          }

\CommentTok{\#\textquotesingle{} TO{-}DO: Explain what this function does in a few sentences}
\CommentTok{\#\textquotesingle{} classify an observation based on the label of the closet
observation in the set of training observation}
```

```
\CommentTok{\#\textquotesingle{} @param xinput          TO{-}DO: explain this argument              A
matrix of featrues for training data observation}
\CommentTok{\#\textquotesingle{} @param y\_binary      TO{-}DO: explain this argument
The vector of training data}
\CommentTok{\#\textquotesingle{} @param xtest          TO{-}DO: explain this argument              A
test observation as a row vector}
\CommentTok{\#\textquotesingle{} @param d             TO{-}DO: explain this argument
A distance function which takes the input to different row vectors.}
\CommentTok{\#\textquotesingle{} @return              TO{-}DO: explain this argument
The predicted label for the test observation}
\NormalTok{nn\_algorithm\_predict }\OtherTok{=} \ControlFlowTok{function}\NormalTok{(xinput,
y\_binary, xtest, }\AttributeTok{d =} \ControlFlowTok{function}\NormalTok{(v1,
v2)\{}\FunctionTok{sum}\NormalTok{((v1 }\SpecialCharTok{{-}}\NormalTok{ v2)}\SpecialCharTok{\^{}}
\DecValTok{2}\NormalTok{)\})\{}
\NormalTok{    distances }\OtherTok{=}
\FunctionTok{array}\NormalTok{ (}\ConstantTok{NA}\NormalTok{, n)}
\NormalTok{    n }\OtherTok{=} \FunctionTok{norw}\NormalTok{(xinput)}
   \ControlFlowTok{for}\NormalTok{(i }\ControlFlowTok{in}
\DecValTok{1}\SpecialCharTok{:}\NormalTok{n)\{}
\NormalTok{       distances[i] }\OtherTok{=} \FunctionTok{d}\NormalTok{(xiput[i, ], xtest)}
\NormalTok{    \}    }

\NormalTok{    y\_binary[}\FunctionTok{which.min}\NormalTok{(distances)]}

\NormalTok{\}}
\end{Highlighting}
\end{Shaded}
```

For extra credit (unless you're a masters student), add an argument \texttt{k} to the \texttt{nn_algorithm_predict} function and update the implementation so it performs KNN. In the case of a tie, choose $\hat{y}$ randomly. Set the default \texttt{k} to be the square root of the size of $\mathcal{D}$ which is an empirical rule-of-thumb popularized by the ``Pattern Classification'' book by Duda, Hart and Stork (2007). Also, alter the documentation in the appropriate places.

```
\begin{Shaded}
\begin{Highlighting}[]
\CommentTok{\#TO{-}DO for the 650 students but extra credit for undergrads}
\CommentTok{\#\textquotesingle{} TO{-}DO: Provide a name for this function                }
\CommentTok{\#\textquotesingle{} one nearest neighbour classifer                          }

\CommentTok{\#\textquotesingle{} TO{-}DO: Explain what this function does in a few sentences}
\CommentTok{\#\textquotesingle{} classify an observation based on the label of the closet
observation in the set of training observation}

\CommentTok{\#\textquotesingle{} @param xinput          TO{-}DO: explain this argument              A
matrix of featrues for training data observation}
\CommentTok{\#\textquotesingle{} @param y\_binary      TO{-}DO: explain this argument
The vector of training data}
\CommentTok{\#\textquotesingle{} @param xtest          TO{-}DO: explain this argument              A
test observation as a row vector}
\CommentTok{\#\textquotesingle{} @param d             TO{-}DO: explain this argument
A distance function which takes the input to different row vectors.}
\CommentTok{\#\textquotesingle{} @param k             TO{-}DO: explain this argument              }
\CommentTok{\#\textquotesingle{} @return              TO{-}DO: explain this argument
The predicted label for the test observation}
```

```
\NormalTok{nn\_algorithm\_predict }\OtherTok{=} \ControlFlowTok{function}\NormalTok{(xinput,
y\_binary, xtest, }\AttributeTok{d =} \ControlFlowTok{function}\NormalTok{(v1,
v2)\{}\FunctionTok{sum}\NormalTok{((v1 }\SpecialCharTok{{-}}\NormalTok{ v2)}\SpecialCharTok{\^{}}
\DecValTok{2}\NormalTok{)\}, k)\{}
\NormalTok{    distances }\OtherTok{=}
\FunctionTok{array}\NormalTok{ (}\ConstantTok{NA}\NormalTok{, n)}
\NormalTok{    n }\OtherTok{=} \FunctionTok{norw}\NormalTok{(xinput)}
   \ControlFlowTok{for}\NormalTok{(i }\ControlFlowTok{in}
\DecValTok{1}\SpecialCharTok{:}\NormalTok{n)\{}
\NormalTok{        distances[i] }\OtherTok{=} \FunctionTok{d}\NormalTok{(xiput[i, ], xtest)}
\NormalTok{    \}    }

\NormalTok{    y\_binary[}\FunctionTok{which.min}\NormalTok{(distances)]}

\NormalTok{\}}
\end{Highlighting}
\end{Shaded}

\hypertarget{basic-binary-classification-modeling}{%
\subsection{Basic Binary Classification
Modeling}\label{basic-binary-classification-modeling}}

\begin{itemize}
\tightlist
\item
   Load the famous \texttt{iris} data frame into the namespace. Provide a
   summary of the columns using the \texttt{skim} function in package
   \texttt{skimr} and write a few descriptive sentences about the
   distributions using the code below and in English.
\end{itemize}

\begin{Shaded}
\begin{Highlighting}[]
\FunctionTok{data}\NormalTok{(iris)}
\NormalTok{pacman}\SpecialCharTok{::}\FunctionTok{p\_load}\NormalTok{(skimr)}
\FunctionTok{skim}\NormalTok{(iris)}
\end{Highlighting}
\end{Shaded}

\begin{longtable}[]{@{}ll@{}}
\caption{Data summary}\tabularnewline
\toprule
\endhead
Name & iris\tabularnewline
Number of rows & 150\tabularnewline
Number of columns & 5\tabularnewline
\_\_\_\_\_ &\tabularnewline
Column type frequency: &\tabularnewline
factor & 1\tabularnewline
numeric & 4\tabularnewline
\_\_\_\_\_ &\tabularnewline
Group variables & None\tabularnewline
\bottomrule
\end{longtable}

\textbf{Variable type: factor}
```

\begin{longtable}[]{@{}lrrlrl@{}}
\toprule
skim_variable & n_missing & complete_rate & ordered & n_unique &
top_counts\tabularnewline
\midrule
\endhead
Species & 0 & 1 & FALSE & 3 & set: 50, ver: 50, vir: 50\tabularnewline
\bottomrule
\end{longtable}

\textbf{Variable type: numeric}

\begin{longtable}[]{@{}lrrrrrrrrrl@{}}
\toprule
skim_variable & n_missing & complete_rate & mean & sd & p0 & p25 &
p50 & p75 & p100 & hist\tabularnewline
\midrule
\endhead
Sepal.Length & 0 & 1 & 5.84 & 0.83 & 4.3 & 5.1 & 5.80 & 6.4 & 7.9 &
▆▇▇▅▂\tabularnewline
Sepal.Width & 0 & 1 & 3.06 & 0.44 & 2.0 & 2.8 & 3.00 & 3.3 & 4.4 &
▁▃▇▂▁\tabularnewline
Petal.Length & 0 & 1 & 3.76 & 1.77 & 1.0 & 1.6 & 4.35 & 5.1 & 6.9 &
▇▁▃▇▂\tabularnewline
Petal.Width & 0 & 1 & 1.20 & 0.76 & 0.1 & 0.3 & 1.30 & 1.8 & 2.5 &
▇▁▃▅▂\tabularnewline
\bottomrule
\end{longtable}

\begin{Shaded}
\begin{Highlighting}[]
\CommentTok{\#TO{-}DO}
\end{Highlighting}
\end{Shaded}

TO-DO: describe this data

The outcome / label / response is \texttt{Species}. This is what we will
be trying to predict. However, we only care about binary classification
between ``setosa'' and ``versicolor'' for the purposes of this exercise.
Thus the first order of business is to drop one class. Let's drop the
data for the level ``virginica'' from the data frame.

\begin{Shaded}
\begin{Highlighting}[]
\CommentTok{\#TO{-}DO}
\NormalTok{iris }\OtherTok{=}\NormalTok{ iris[iris}\SpecialCharTok{$}\NormalTok{Species }\SpecialCh
arTok{!=} \StringTok{"virginica"}\NormalTok{ ,] }
\end{Highlighting}
\end{Shaded}

Now create a vector \texttt{y} that is length the number of remaining
rows in the data frame whose entries are 0 if ``setosa'' and 1 if
``versicolor''.

\begin{Shaded}

\begin{Highlighting}[]
\NormalTok{y }\OtherTok{=}
\FunctionTok{as.integer}\NormalTok{(iris}\SpecialCharTok{$}\NormalTok{Species }\SpecialCharTok{==}
\StringTok{"setosa"}\NormalTok{)}
\NormalTok{y}
\end{Highlighting}
\end{Shaded}

\begin{verbatim}
[1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
[38] 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
[75] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
\end{verbatim}

\begin{Shaded}
\begin{Highlighting}[]
\CommentTok{\#TO{-}DO}
\end{Highlighting}
\end{Shaded}

\begin{itemize}
\tightlist
\item
  Write a function \texttt{mode} returning the sample mode.
\end{itemize}

\begin{Shaded}
\begin{Highlighting}[]
\NormalTok{mode }\OtherTok{=} \ControlFlowTok{function}\NormalTok{(v)\{}

\FunctionTok{names}\NormalTok{(}\FunctionTok{sort}\NormalTok{(}\FunctionTok{table}\NormalTok{(
v), }\AttributeTok{decreasing =} \ConstantTok{TRUE}\NormalTok{)[}\DecValTok{1}\NormalTok{])}

  \CommentTok{\# table(v)}
  \CommentTok{\# table(sample(letters, 1000, replace = TRUE))}
  \CommentTok{\# sort(table(sample(letters, 1000, replace = TRUE)), decreasing = TURE)}
  \CommentTok{\# sorted_count = sort(table(sample(letters, 1000, replace = TRUE)), decreasing =
TURE)}
\NormalTok{\}}
\FunctionTok{mode}\NormalTok{(v)}
\end{Highlighting}
\end{Shaded}

\begin{verbatim}
[1] "1"
\end{verbatim}

\begin{itemize}
\tightlist
\item
  Fit a threshold model to \texttt{y} using the feature
  \texttt{Sepal.Length}. Write your own code to do this. What is the
  estimated value of the threshold parameter? Save the threshold value
  as \texttt{threshold}.
\end{itemize}

\begin{Shaded}

```
\begin{Highlighting}[]
\CommentTok{\#TO{-}DO}
\end{Highlighting}
\end{Shaded}
```

What is the total number of errors this model makes?

```
\begin{Shaded}
\begin{Highlighting}[]
\CommentTok{\#TO{-}DO}
\end{Highlighting}
\end{Shaded}
```

Does the threshold model's performance make sense given the following summaries:

```
\begin{Shaded}
\begin{Highlighting}[]
\CommentTok{\#threshold}
\FunctionTok{summary}\NormalTok{(iris[iris}\SpecialCharTok{$}\NormalTok{Species }\SpecialCharTok
{==} \StringTok{"setosa"}\NormalTok{, }\StringTok{"Sepal.Length"}\NormalTok{])}
\end{Highlighting}
\end{Shaded}
```

```
\begin{verbatim}
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   4.300   4.800   5.000   5.006   5.200   5.800
\end{verbatim}
```

```
\begin{Shaded}
\begin{Highlighting}[]
\FunctionTok{summary}\NormalTok{(iris[iris}\SpecialCharTok{$}\NormalTok{Species }\SpecialCharTok
{==} \StringTok{"versicolor"}\NormalTok{, }\StringTok{"Sepal.Length"}\NormalTok{])}
\end{Highlighting}
\end{Shaded}
```

```
\begin{verbatim}
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   4.900   5.600   5.900   5.936   6.300   7.000
\end{verbatim}
```

TO-DO: Write your answer here in English.

Create the function \texttt{g} explicitly that can predict \texttt{y} from \texttt{x} being a new \texttt{Sepal.Length}.

```
\begin{Shaded}
\begin{Highlighting}[]
\NormalTok{g }\OtherTok{=} \ControlFlowTok{function}\NormalTok{(x)\{}
  \CommentTok{\#TO{-}DO}
\NormalTok{\}}
\end{Highlighting}
\end{Shaded}
```

```
\hypertarget{perceptron}{%
\subsection{Perceptron}\label{perceptron}}
```

You will code the ``perceptron learning algorithm'' for arbitrary number
of features \(p\). Take a look at the comments above the function.
Respect the spec below:

```
\begin{Shaded}
\begin{Highlighting}[]
\CommentTok{\#\textquotesingle{} TO{-}DO: Provide a name for this function }
\CommentTok{\#\textquotesingle{}}
\CommentTok{\#\textquotesingle{} TO{-}DO: Explain what this function does in a few sentences}
\CommentTok{\#\textquotesingle{}}
\CommentTok{\#\textquotesingle{} @param Xinput           TO{-}DO: Explain this}
\CommentTok{\#\textquotesingle{} @param y\_binary       TO{-}DO: Explain this}
\CommentTok{\#\textquotesingle{} @param MAX\_ITER       TO{-}DO: Explain this}
\CommentTok{\#\textquotesingle{} @param w               TO{-}DO: Explain this}
\CommentTok{\#\textquotesingle{}}
\CommentTok{\#\textquotesingle{} @return                The computed final parameter (weight) as
a vector of length p + 1}
\NormalTok{perceptron\_learning\_algorithm }\OtherTok{=}
\ControlFlowTok{function}\NormalTok{(Xinput, y\_binary, }\AttributeTok{MAX\_ITER =}
\DecValTok{1000}\NormalTok{, }\AttributeTok{w =} \ConstantTok{NULL}\NormalTok{)\{}
  \CommentTok{\#TO{-}DO}
\NormalTok{\}}
\end{Highlighting}
\end{Shaded}
```

To understand what the algorithm is doing - linear ``discrimination''
between two response categories, we can draw a picture. First let's make
up some very simple training data \(\mathbb{D}\).

```
\begin{Shaded}
\begin{Highlighting}[]
\NormalTok{Xy\_simple }\OtherTok{=} \FunctionTok{data.frame}\NormalTok{(}
  \AttributeTok{response =}
\FunctionTok{factor}\NormalTok{(}\FunctionTok{c}\NormalTok{(}\DecValTok{0}\NormalTok{, }\DecVa
lTok{0}\NormalTok{, }\DecValTok{0}\NormalTok{, }\DecValTok{1}\NormalTok{, }\DecValTok{1}\Norma
lTok{, }\DecValTok{1}\NormalTok{)), }\CommentTok{\#nominal}
  \AttributeTok{first\_feature =}
\FunctionTok{c}\NormalTok{(}\DecValTok{1}\NormalTok{, }\DecValTok{1}\NormalTok{, }\DecValTok{2}
\NormalTok{, }\DecValTok{3}\NormalTok{, }\DecValTok{3}\NormalTok{, }\DecValTok{4}\NormalTok{),
       }\CommentTok{\#continuous}
  \AttributeTok{second\_feature =}
\FunctionTok{c}\NormalTok{(}\DecValTok{1}\NormalTok{, }\DecValTok{2}\NormalTok{, }\DecValTok{1}
\NormalTok{, }\DecValTok{3}\NormalTok{, }\DecValTok{4}\NormalTok{, }\DecValTok{3}\NormalTok{)
     }\CommentTok{\#continuous}
\NormalTok{)}
\end{Highlighting}
\end{Shaded}
```

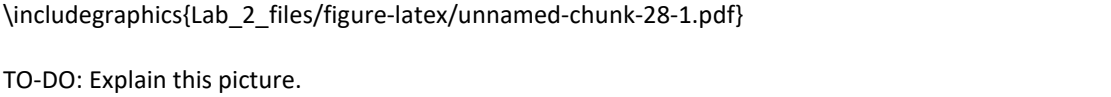
We haven't spoken about visualization yet, but it is important we do
some of it now. Thus, I will write this code for you and you will just
run it. First we load the visualization library we're going to use:

```
\begin{Shaded}
\begin{Highlighting}[]
\NormalTok{pacman}\SpecialCharTok{::}\FunctionTok{p\_load}\NormalTok{(ggplot2)}
\end{Highlighting}
\end{Shaded}
```

We are going to just get some plots and not talk about the code to generate them as we will have a whole unit on visualization using \texttt{ggplot2} in the future.

Let's first plot \(y\) by the two features so the coordinate plane will be the two features and we use different colors to represent the third dimension, \(y\).

```
\begin{Shaded}
\begin{Highlighting}[]
\NormalTok{simple\_viz\_obj }\OtherTok{=}
\FunctionTok{ggplot}\NormalTok{(Xy\_simple, }\FunctionTok{aes}\NormalTok{(}\AttributeTok{x
=}\NormalTok{ first\_feature, }\AttributeTok{y =}\NormalTok{ second\_feature, }\AttributeTok{color
=}\NormalTok{ response)) }\SpecialCharTok{+}
   \FunctionTok{geom\_point}\NormalTok{(}\AttributeTok{size =} \DecValTok{5}\NormalTok{)}
\NormalTok{simple\_viz\_obj}
\end{Highlighting}
\end{Shaded}
```



TO-DO: Explain this picture.

Now, let us run the algorithm and see what happens:

```
\begin{Shaded}
\begin{Highlighting}[]
\NormalTok{w\_vec\_simple\_per }\OtherTok{=}
\FunctionTok{perceptron\_learning\_algorithm}\NormalTok{(}
   \FunctionTok{cbind}\NormalTok{(Xy\_simple}\SpecialCharTok{$}\NormalTok{first\_feature,
Xy\_simple}\SpecialCharTok{$}\NormalTok{second\_feature),}

\FunctionTok{as.numeric}\NormalTok{(Xy\_simple}\SpecialCharTok{$}\NormalTok{response }\Special
CharTok{==} \DecValTok{1}\NormalTok{))}
\NormalTok{w\_vec\_simple\_per}
\end{Highlighting}
\end{Shaded}
```

```
\begin{verbatim}
## NULL
\end{verbatim}
```

Explain this output. What do the numbers mean? What is the intercept of this line and the slope? You will have to do some algebra.

TO-DO

```
\begin{Shaded}
\begin{Highlighting}[]
\CommentTok{\#simple\_perceptron\_line = geom\_abline()
\CommentTok{\#    intercept = {-}w\_vec\_simple\_per[1] / w\_vec\_simple\_per[3], }
\CommentTok{\#    slope = {-}w\_vec\_simple\_per[2] / w\_vec\_simple\_per[3], }
\CommentTok{\#    color = "orange")}
\CommentTok{\# simple\_viz\_obj + simple\_perceptron\_line}
\end{Highlighting}
\end{Shaded}
```

Explain this picture. Why is this line of separation not ``satisfying''
to you?

TO-DO

For extra credit, program the maximum-margin hyperplane perceptron that
provides the best linear discrimination model for linearly separable
data. Make sure you provide ROxygen documentation for this function.

\begin{Shaded}
\begin{Highlighting}[]
\CommentTok{\#TO{-}DO}
\end{Highlighting}
\end{Shaded}

\end{document}