Work the following problems from Christensen (C) and Wakefield (W)

1. 5.8 (W) (see link to eBook on Calendar)

2. 1.5.8 (C) (see link to eBook on Calendar)

3. We showed that $\mathsf{P}_{\mathbf{X}} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ was an orthogonal projection on the column space of $\mathbf{X}$ and that $\hat{\mathbf{Y}} = P_{\mathbf{X}}\mathbf{Y}$. While useful for theory, the projection matrix should never be used in practice to find the MLE of $\boldsymbol{\mu}$ due to 1) computational complexity (inverses and matrix multiplication) and instability. To find $\hat{\boldsymbol{\beta}}$ we solve $\mathbf{X}\boldsymbol{\beta} = \mathsf{P}_{\mathbf{X}}\mathbf{Y}$ which leads to the *normal equations* $(\mathbf{X}^T\mathbf{X})\boldsymbol{\beta} = \mathbf{X}^T\mathbf{Y}$ and solving the system of equations for $\boldsymbol{\beta}$. Instead consider the following for $\mathbf{X}$ $(n \times p, p < n)$ of rank p

   (a) Any $\mathbf{X}$ may be written via a singular value decomposition as $\mathbf{U}\Lambda\mathbf{V}^T$ where $\mathbf{U}$ is a $n \times p$ orthonormal matrix ($\mathbf{U}^T\mathbf{U} = \mathbf{I}_p$ and columns of $\mathbf{U}$ form an orthonormal basis (ONB) for $C(\mathbf{X})$), Λ is a $p \times p$ diagonal matrix and $\mathbf{V}$ is a $p \times p$ orthogonal matrix ($\mathbf{V}^T\mathbf{V} = \mathbf{V}\mathbf{V}^T = \mathbf{I}_p$. Note the difference between *orthonormal* and *orthogonal*. Show that P_X may be expressed as a function of $\mathbf{U}$ only and provide an expression for $\hat{\mathbf{Y}}$. Similarly, find an expression for $hat\boldsymbol{\beta}$ in terms of $\mathbf{U}$, Λ and $\mathbf{V}$. Your result should only require the inverse of a diagonal matrix!

   (b) $\mathbf{X}$ may be written in a (reduced or thinned) QR decomposition as a matrix $\mathbf{Q}$ that is a $n \times p$ orthonormal matrix (which forms an ONB for $C(\mathbf{X})$) and $\mathbf{R}$ which is a $p \times p$ upper triangular matrix (i.e all elements below the diagonal are 0) where $\mathbf{X} = \mathbf{QR}$. The columns of $\mathbf{Q}$ are an ONB for the $C(\mathbf{X})$. Show that $\mathsf{P}_{\mathbf{X}}$ may be expressed as a function of $\mathbf{Q}$ alone. Show that the the normal equations reduce to solving the triangular system $\mathbf{R}\boldsymbol{\beta} = \mathbf{Z}$ where $\mathbf{Z} = \mathbf{Q}^T\mathbf{Y}$. Because $\mathbf{R}$ is upper triangular, show that $\hat{\boldsymbol{\beta}}$ may be obtained be back-solving (and avoiding the matrix inverse of $\mathbf{X}^T\mathbf{X}$.

   (c) Any symmetric matrix $\mathbf{A}$ may be written via a Cholesky decomposition as $\mathbf{A} = \mathbf{L}\mathbf{L}^T$ where $\mathbf{L}$ is lower triangular. If $\mathbf{Z} = \mathbf{X}^T\mathbf{Y}$ show that we can solve two triangular systems $\mathbf{L}\mathbf{L}^T\boldsymbol{\beta} = \mathbf{Z}$ by solving for $\mathbf{w}$ using $\mathbf{L}\mathbf{w} = \mathbf{Z}$ using a forward substitution and then for $\hat{\boldsymbol{\beta}}$ using $\mathbf{L}^T\boldsymbol{\beta} = \mathbf{w}$ avoiding any matrix inversion.

   (d) Use R to find $\mathbf{Q}$ and $\mathbf{U}$ for the matrices in problems 1.5.8 in Christensen. Does $\mathbf{Q}$ equal $\mathbf{U}$? (see `help(qr)` and `help(svd)`.

   (e) Prove that the two projection matrices obtained by the SVD and the QR method are the same. (Hint: review Theorems in Christensen Appencicies about uniqueness of projections)

   Note: The Cholesky method is the fastest in terms of $O(np^2 + p^3/3)$ floating point operations (flops), but is numerically unstable if the matrix is poorly conditioned. R uses the QR method $(O(2np^2 - 2p^3/3)$ flops in the function `lm.fit()` (which is the workhorse underneath the `lm()` function. Generalized QR algorithms can handle rank deficient case. The SVD method is the most expensive $O(2np^2 + 11n^3)$ but can handle the rank case. There are generalized Cholesky and QR methods for the rank deficient case.