Exercise 1

(a)

$$
\begin{aligned}
\frac{1}{N}\sum_{i=1}^{N}\|\mathbf{x}_i-\bar{\mathbf{x}}\|^2 &= \frac{1}{N}\sum_{i=1}^{N}(\mathbf{x}_i-\bar{\mathbf{x}})^\top(\mathbf{x}_i-\bar{\mathbf{x}}) \\
&= \frac{1}{N}\sum_{i=1}^{N}\operatorname{tr}\Big((\mathbf{x}_i-\bar{\mathbf{x}})^\top(\mathbf{x}_i-\bar{\mathbf{x}})\Big) \\
&= \frac{1}{N}\sum_{i=1}^{N}\operatorname{tr}\Big((\mathbf{x}_i-\bar{\mathbf{x}})(\mathbf{x}_i-\bar{\mathbf{x}})^\top\Big) \\
&= \operatorname{tr}\left(\frac{1}{N}\sum_{i=1}^{N}(\mathbf{x}_i-\bar{\mathbf{x}})(\mathbf{x}_i-\bar{\mathbf{x}})^\top\right) \\
&= \operatorname{tr}(\mathbf{\Sigma})
\end{aligned}
$$

(b)

$$
\frac{1}{N}\sum_{i=1}^{N}\|\mathbf{x}_i\|^2 = \frac{1}{N}\sum_{i=1}^{N}\|\mathbf{x}_i-\bar{\mathbf{x}}\|^2 = \operatorname{tr}(\mathbf{\Sigma}) = \sum_{k=1}^{p}\operatorname{Var}(X_k) = \sum_{k=1}^{p}1 = p
$$

Exercise 2

(a) In multiple linear regression we have the model

$$
Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p + \varepsilon
$$

Note that if we take the expectation from both sites, we get

$$
\mathbb{E}[Y] = \beta_0 + \beta_1\mathbb{E}[X_1] + \cdots + \beta_p\mathbb{E}[X_p] + \mathbb{E}[\varepsilon]
$$

Since the predictors and observations have zero-mean, then $\beta_0 = 0$.

(b) The SVD of the data matrix is $\mathbf{X} = \mathbf{UDV}^\top$ so if we plug this into the expression for $\hat{\beta}$ we get

$$
\hat{\beta} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y} = (\mathbf{VDU}^\top\mathbf{UDV}^\top)^{-1}\mathbf{VDU}^\top\mathbf{y} = (\mathbf{VD}^2\mathbf{V}^\top)^{-1}\mathbf{VDU}^\top\mathbf{y} = \mathbf{VD}^{-2}\mathbf{V}^\top\mathbf{VDU}^\top\mathbf{y}
$$

so in the end we get $\hat{\beta} = \mathbf{VD}^{-1}\mathbf{U}^\top\mathbf{y}$ and $\hat{\beta}_i = \displaystyle\sum_{k=1}^{p}\frac{\mathbf{u}_i^\top\mathbf{y}}{d_k}\mathbf{v}_{ik}$

Note also that the predictions with the model are $\hat{\mathbf{y}} = \mathbf{X}\hat{\beta} = \mathbf{UDV}^\top\mathbf{VD}^{-1}\mathbf{U}^\top\mathbf{y} = \mathbf{UU}^\top\mathbf{y}$

(c) Matrix $\mathbf{Z}$ is the projection of the data matrix on its q-top principal components. Therefore, we have:

$$\underbrace{\mathbf{Z}}_{n\times q} = \underbrace{\mathbf{X}}_{n\times p}\underbrace{\mathbf{V}_q}_{p\times q} = \underbrace{\mathbf{U}}_{n\times p}\underbrace{\mathbf{D}}_{p\times p}\underbrace{\mathbf{V}^\top}_{p\times p}\underbrace{\mathbf{V}_q}_{p\times q} = \mathbf{U}\mathbf{D}\underbrace{\begin{bmatrix}\mathbf{I}_q \\ \mathbf{0}_{(p-q)\times q}\end{bmatrix}}_{p\times q} = \underbrace{\mathbf{U}}_{n\times p}\underbrace{\mathbf{D}_q}_{p\times q} \quad \text{where} \quad \mathbf{D}_q = \begin{bmatrix} d_1 & & \\ & \ddots & \\ & & d_q \\ & \mathbf{0}_{(p-q)\times q} & \end{bmatrix}$$

We calculate the coefficients for the new regression model

$$\hat{\gamma} = (\mathbf{Z}^\top\mathbf{Z})^{-1}\mathbf{Z}^\top\mathbf{y} = (\mathbf{D}_q^\top\mathbf{U}^\top\mathbf{U}\mathbf{D}_q)^{-1}\mathbf{D}_q\mathbf{U}^\top\mathbf{y} = \underbrace{(\mathbf{D}_q^\top\mathbf{D}_q)^{-1}}_{q\times q}\underbrace{\mathbf{D}_q^\top}_{q\times p}\underbrace{\mathbf{U}^\top}_{p\times n}\underbrace{\mathbf{y}}_{n\times 1} = \begin{bmatrix}\frac{1}{d_1}\mathbf{u}_1^\top\mathbf{y} \\ \vdots \\ \frac{1}{d_q}\mathbf{u}_q^\top\mathbf{y}\end{bmatrix}$$

where we note that the γ coefficients can be calculated as if we had q independent simple linear regressions. This is due to the diagonal shape of the matrix $\mathbf{Z}$ as per:

$$\mathbf{Z}^\top\mathbf{Z} = \mathbf{D}_q^\top\mathbf{D}_q = \begin{bmatrix} d_1 & & \\ & \ddots & \\ & & d_q \end{bmatrix}$$

If we take it back to the original space, we get

$$\hat{\beta}^{\text{PCR}} = \mathbf{V}_q\hat{\gamma} = \mathbf{V}_q\begin{bmatrix}\frac{1}{d_1}\mathbf{u}_1^\top\mathbf{y} \\ \vdots \\ \frac{1}{d_q}\mathbf{u}_q^\top\mathbf{y}\end{bmatrix} = [\mathbf{v}_1 \quad \dots \quad \mathbf{v}_q]\begin{bmatrix}\frac{1}{d_1}\mathbf{u}_1^\top\mathbf{y} \\ \vdots \\ \frac{1}{d_q}\mathbf{u}_q^\top\mathbf{y}\end{bmatrix} \quad \text{and} \quad \hat{\beta}_i^{\text{PCR}} = \sum_{k=1}^{q}\frac{\mathbf{u}_i^\top\mathbf{y}}{d_k}\mathbf{v}_{ik}$$

(d) We notice that the parameters for the linear regression obtained with the q-top principal components is a truncated version of the original least squares parameters. We observe that the terms of the sum depending of small singular values have been discarded.