Matrices - Storing Tables of Data

```r
matrix(2, 3, 2)
```

```
     [,1] [,2]
[1,]    2    2
[2,]    2    2
[3,]    2    2
```

```r
matrix(c(1, -1, 2, 3, 2, -2), nrow = 2, ncol = 3,
    byrow = TRUE)
```

```
     [,1] [,2] [,3]
[1,]    1   -1    2
[2,]    3    2   -2
```

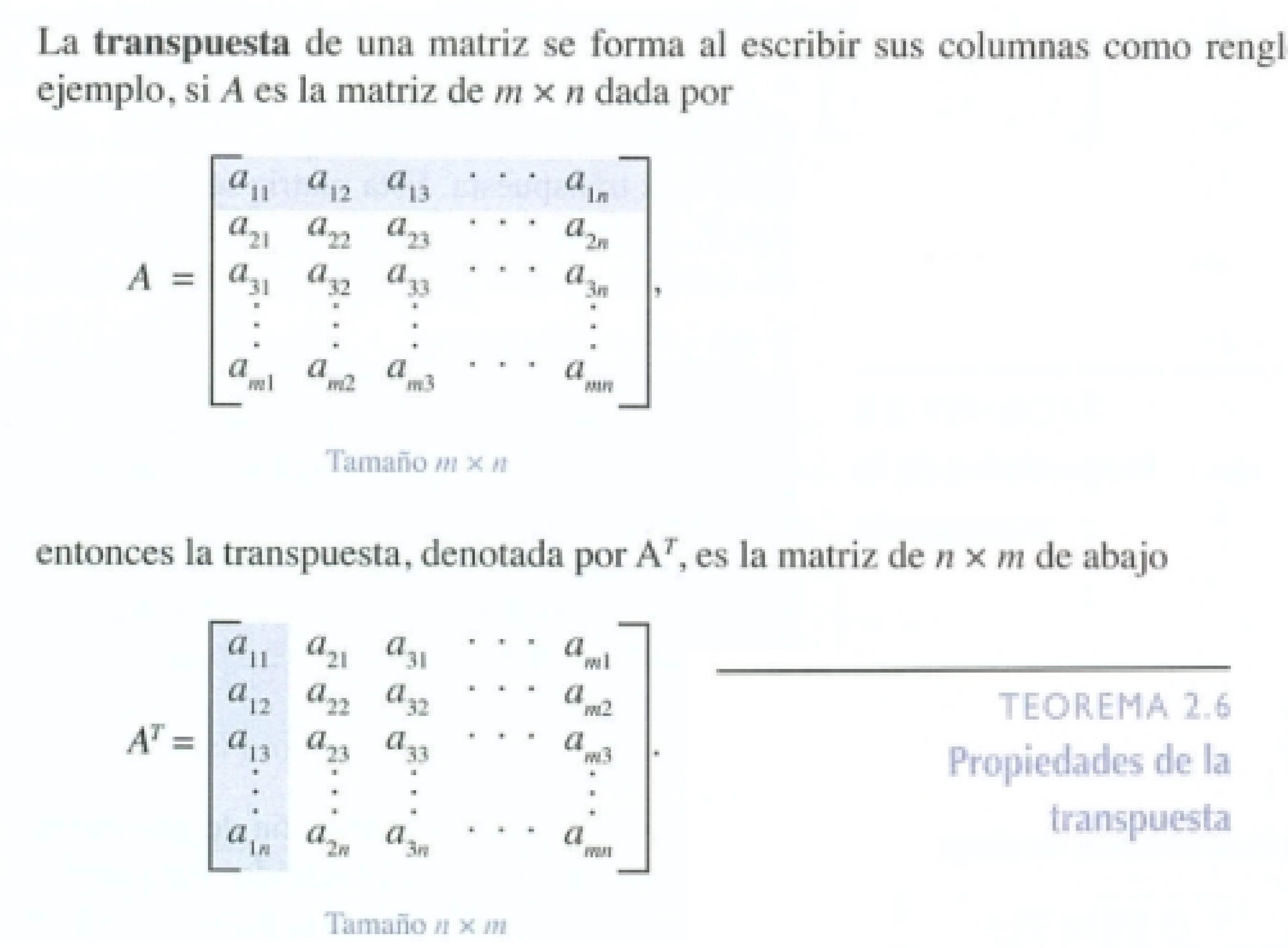
```r
matrix(c(1, -1, 2, 3, 2, -2), nrow = 2, ncol = 3,
    byrow = FALSE)
```

```
     [,1] [,2] [,3]
[1,]    1    2    2
[2,]   -1    3   -2
```

```r
A[2, 1] <- 100
print(A)
```

```
     [,1] [,2] [,3]
[1,]    1    2    2
[2,]  100    3   -2
```

Transpuesta de una matriz



La **transpuesta** de una matriz se forma al escribir sus columnas como renglones. Por ejemplo, si A es la matriz de $m \times n$ dada por

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mn} \end{bmatrix},$$

Tamaño $m \times n$

entonces la transpuesta, denotada por A^T, es la matriz de $n \times m$ de abajo

$$A^T = \begin{bmatrix} a_{11} & a_{21} & a_{31} & \cdots & a_{m1} \\ a_{12} & a_{22} & a_{32} & \cdots & a_{m2} \\ a_{13} & a_{23} & a_{33} & \cdots & a_{m3} \\ \vdots & \vdots & \vdots & & \vdots \\ a_{1n} & a_{2n} & a_{3n} & \cdots & a_{mn} \end{bmatrix}.$$

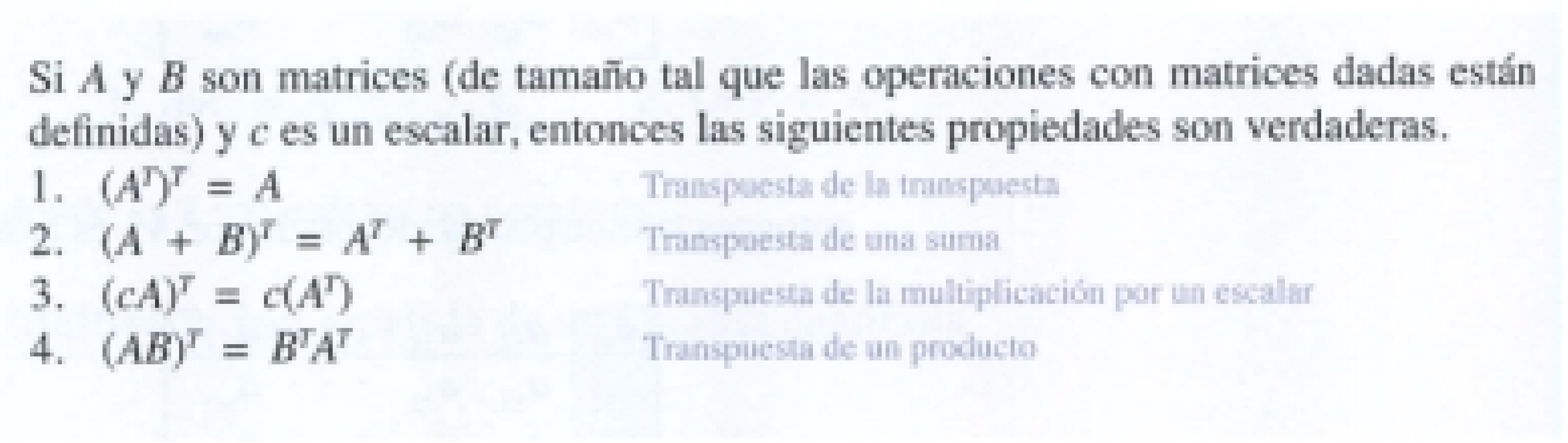
Tamaño $n \times m$

$$\vec{x} = \begin{pmatrix} 1 \\ 5 \\ -2 \\ 4 \end{pmatrix} \qquad \vec{x}^T = \begin{pmatrix} 1 & 5 & -2 & 4 \end{pmatrix}$$

$$\vec{y} = \begin{pmatrix} 11 & -7 & 12 & 14 & 21 \end{pmatrix} \qquad \vec{y}^T = \begin{pmatrix} 11 \\ -7 \\ 12 \\ 14 \\ 21 \end{pmatrix}$$

TEOREMA 2.6

Propiedades de la transpuesta



Si A y B son matrices (de tamaño tal que las operaciones con matrices dadas están definidas) y c es un escalar, entonces las siguientes propiedades son verdaderas.

1. $(A^T)^T = A$ Transpuesta de la transpuesta
2. $(A + B)^T = A^T + B^T$ Transpuesta de una suma
3. $(cA)^T = c(A^T)$ Transpuesta de la multiplicación por un escalar
4. $(AB)^T = B^T A^T$ Transpuesta de un producto

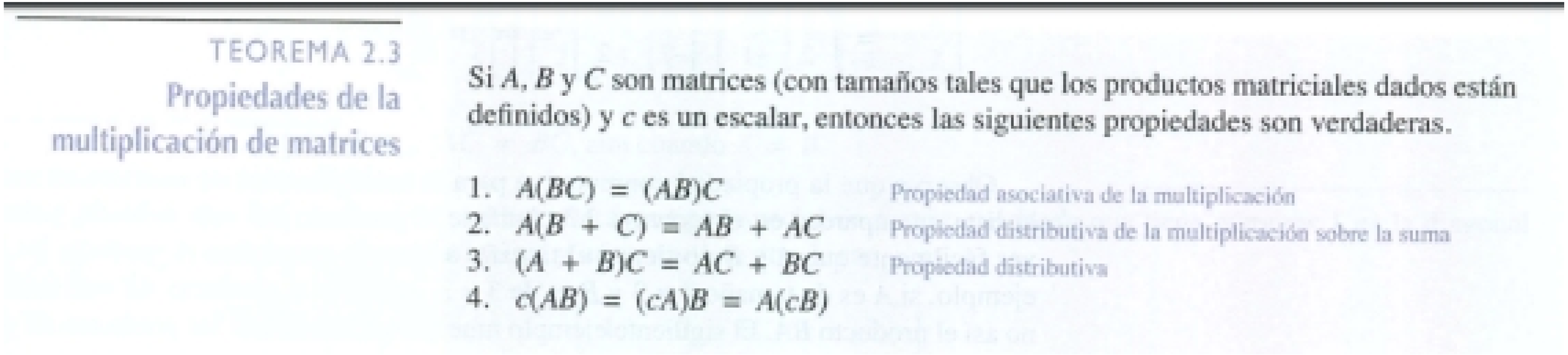
Matrix-Vector Multiplication

$$A\vec{x} =$$

$$\underbrace{\begin{pmatrix} 1 & -1 \\ 2 & 1 \\ 4 & -2 \end{pmatrix}}_{3x2} \times \underbrace{\begin{pmatrix} 1 \\ 2 \end{pmatrix}}_{2x1} = \begin{pmatrix} 1\times 1 & + & (-1)\times 2 \\ 2\times 1 & + & 1\times 2 \\ 4\times 1 & + & (-2)\times 2 \end{pmatrix} = \underbrace{\begin{pmatrix} -1 \\ 4 \\ 0 \end{pmatrix}}_{3x1}$$

```
A%*%b
```

```
      [,1]
[1,]     7
[2,]     9
```



TEOREMA 2.3

Propiedades de la multiplicación de matrices

Si A, B y C son matrices (con tamaños tales que los productos matriciales dados están definidos) y c es un escalar, entonces las siguientes propiedades son verdaderas.

1. $A(BC) = (AB)C$ — Propiedad asociativa de la multiplicación
2. $A(B + C) = AB + AC$ — Propiedad distributiva de la multiplicación sobre la suma
3. $(A + B)C = AC + BC$ — Propiedad distributiva
4. $c(AB) = (cA)B = A(cB)$

Motivation - Can These Vectors Make That Vector?

$$= \begin{pmatrix} 1 \times 1 & + & (-1) \times 2 \\ 2 \times 1 & + & 1 \times 2 \\ 4 \times 1 & + & (-2) \times 2 \end{pmatrix}$$

$$= 1 \times \begin{pmatrix} 1 \\ 2 \\ 4 \end{pmatrix} + 2 \times \begin{pmatrix} -1 \\ 1 \\ -2 \end{pmatrix}$$

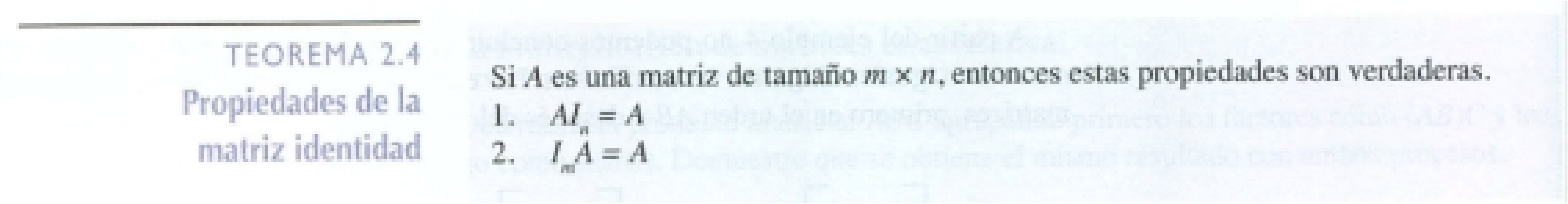
The Identity Matrix

```
I <- diag(2)
print(I)
```

```
     [,1] [,2]
[1,]    1    0
[2,]    0    1
```

```
A%*%I
```

```
     [,1] [,2]
[1,]    1    2
[2,]    2    1
```



TEOREMA 2.4

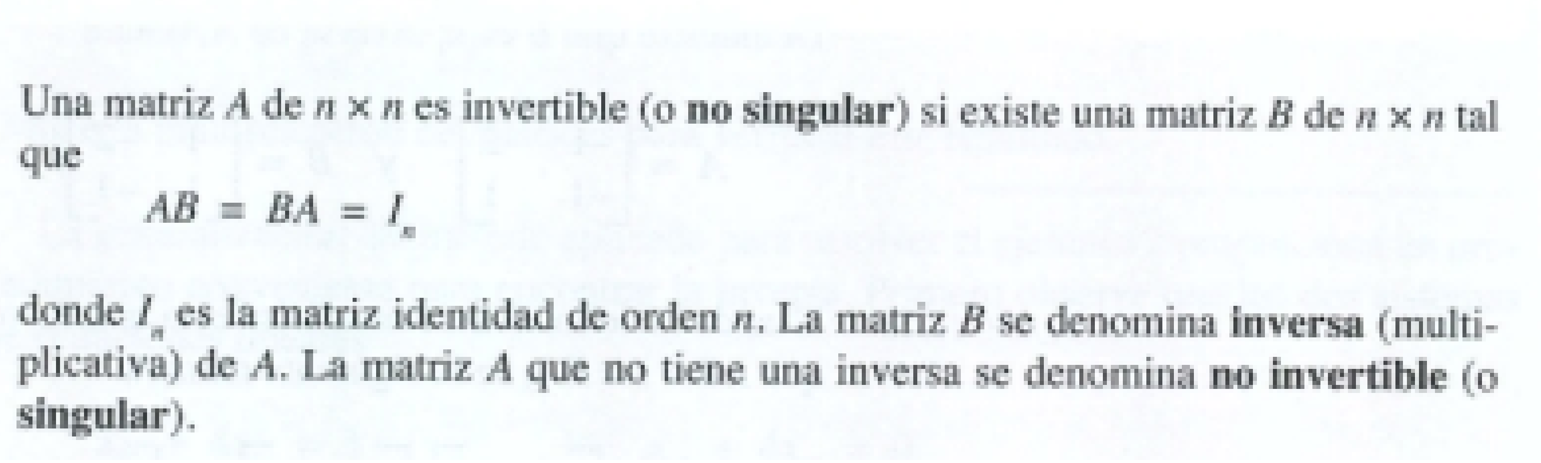
Propiedades de la matriz identidad

Si A es una matriz de tamaño $m \times n$, entonces estas propiedades son verdaderas.

1. $AI_n = A$
2. $I_mA = A$

The Inverse Matrix

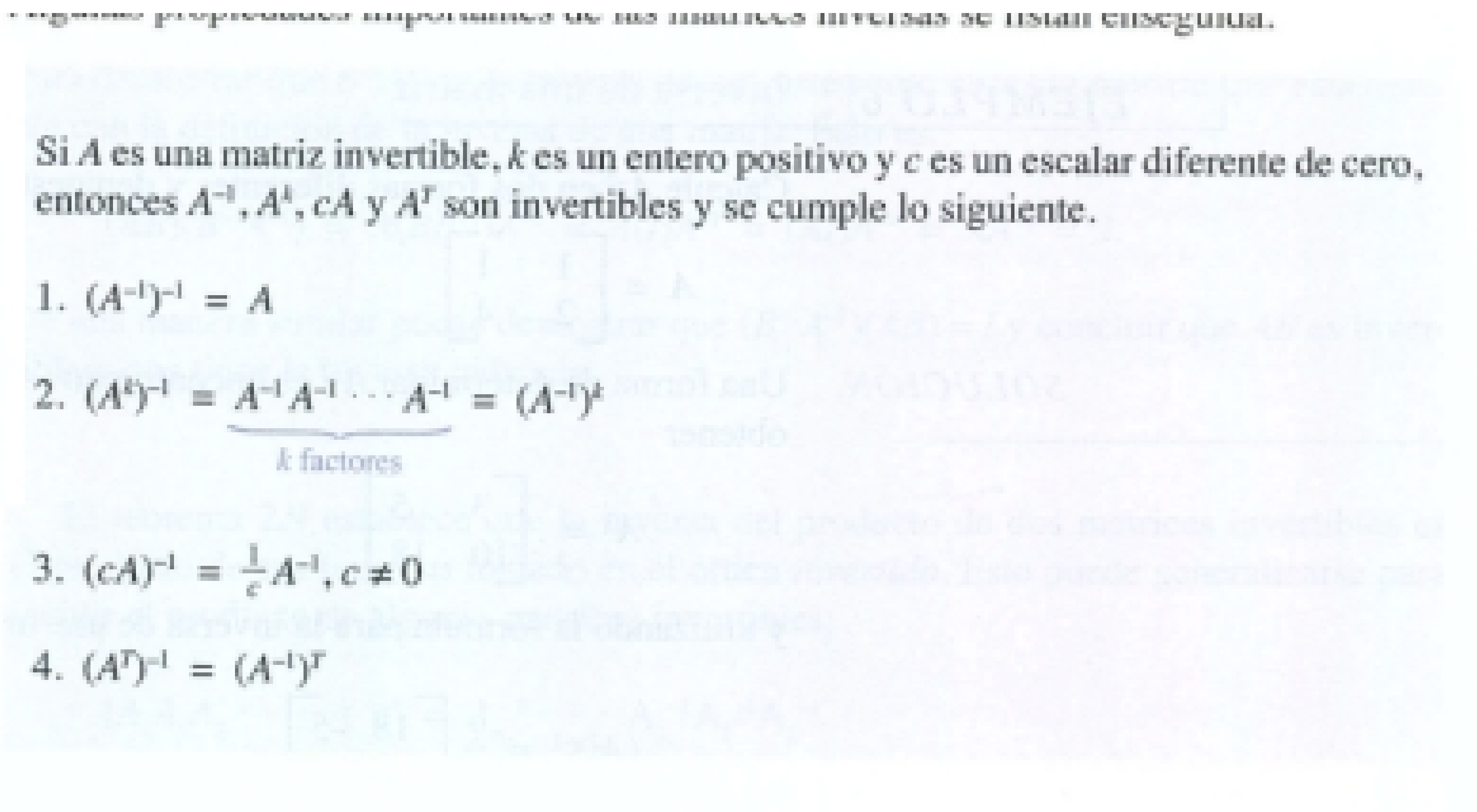
The **solve()** function in R will find the inverse of a matrix if it exists and provide an error if it does not.



Una matriz A de $n \times n$ es invertible (o **no singular**) si existe una matriz B de $n \times n$ tal que

$$AB = BA = I_n$$

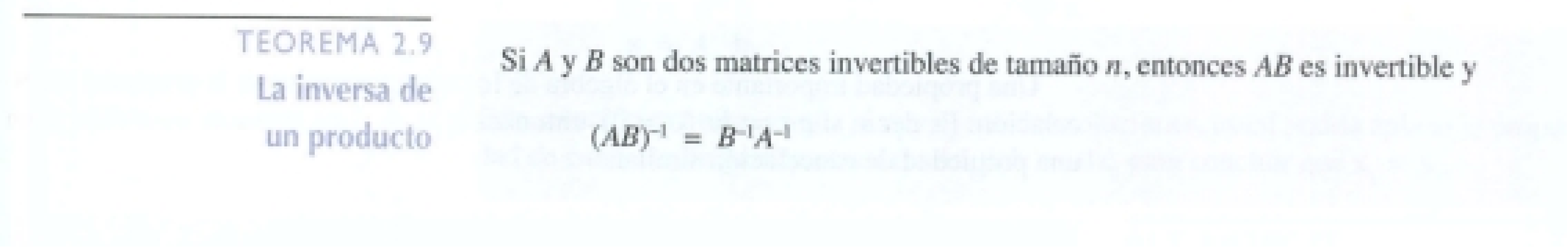
donde I_n es la matriz identidad de orden n. La matriz B se denomina **inversa** (multiplicativa) de A. La matriz A que no tiene una inversa se denomina **no invertible** (o **singular**).



Algunas propiedades importantes de las matrices inversas se listan enseguida.

Si A es una matriz invertible, k es un entero positivo y c es un escalar diferente de cero, entonces A^{-1}, A^k, cA y A^T son invertibles y se cumple lo siguiente.

1. $(A^{-1})^{-1} = A$
2. $(A^k)^{-1} = \underbrace{A^{-1}A^{-1} \cdots A^{-1}}_{k \text{ factores}} = (A^{-1})^k$
3. $(cA)^{-1} = \frac{1}{c}A^{-1}, c \neq 0$
4. $(A^T)^{-1} = (A^{-1})^T$



TEOREMA 2.9
La inversa de un producto

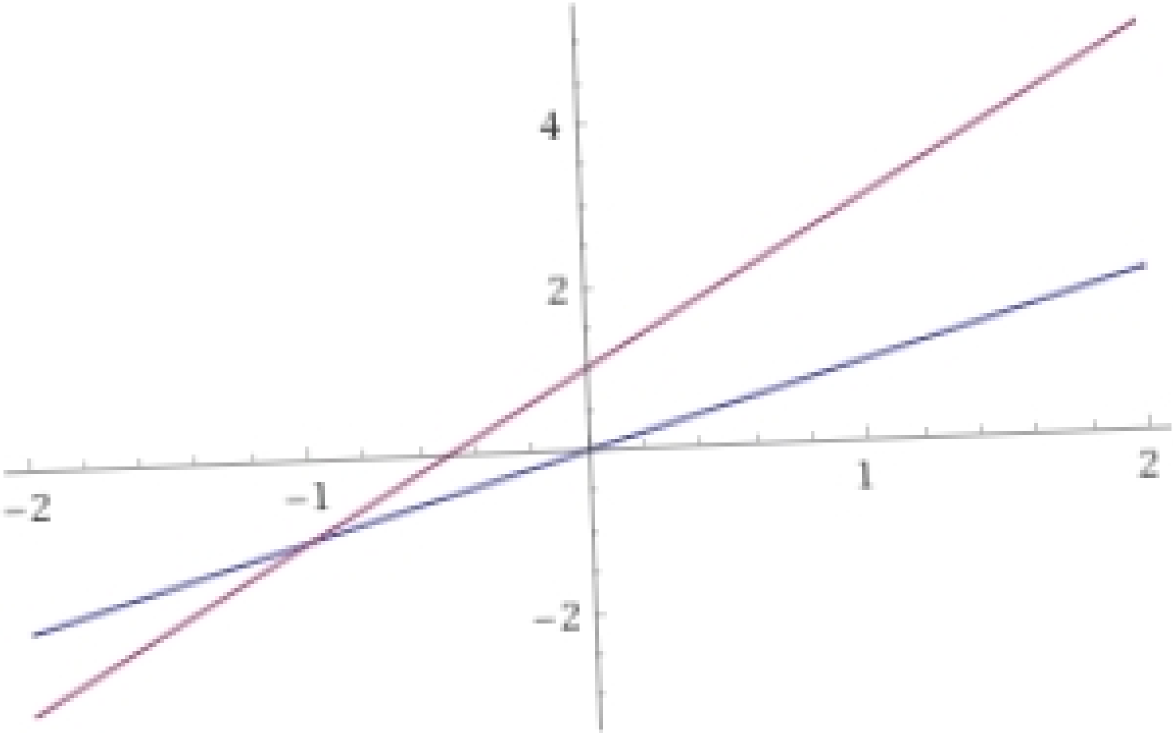
Si A y B son dos matrices invertibles de tamaño n, entonces AB es invertible y

$$(AB)^{-1} = B^{-1}A^{-1}$$

Conditions for a Unique Solution to Matrix-Vector Equations

If A is an n by n square matrix, then the following conditions are equivalent and imply a unique solution to

$$Ax = b$$



- The matrix A has an inverse (is invertible)
- The determinant of A is nonzero
- The rows and columns of A form a basis for the set of all vectors with n elements
- The homogeneous equation $Ax = 0$ has just the trivial ($x = 0$) solution

Motivation - Can These Vectors Make That Vector?

$$A\vec{x} = \vec{b}$$

$$\begin{pmatrix} 4 & -2 \\ -3 & 2 \end{pmatrix} \times \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

$$x_1 \times \begin{pmatrix} 4 \\ -3 \end{pmatrix} + x_2 \times \begin{pmatrix} -2 \\ 2 \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

$$\begin{aligned} A\vec{x} &= \vec{b} \\ A^{-1}A\vec{x} &= A^{-1}\vec{b} \\ I\vec{x} &= A^{-1}\vec{b} \\ \vec{x} &= A^{-1}\vec{b} \end{aligned}$$

Solving Matrix-Vector Equations

```
print(A)
```

```
     [,1] [,2]
[1,]    1   -2
[2,]    0    4
```

```
print(b)
```

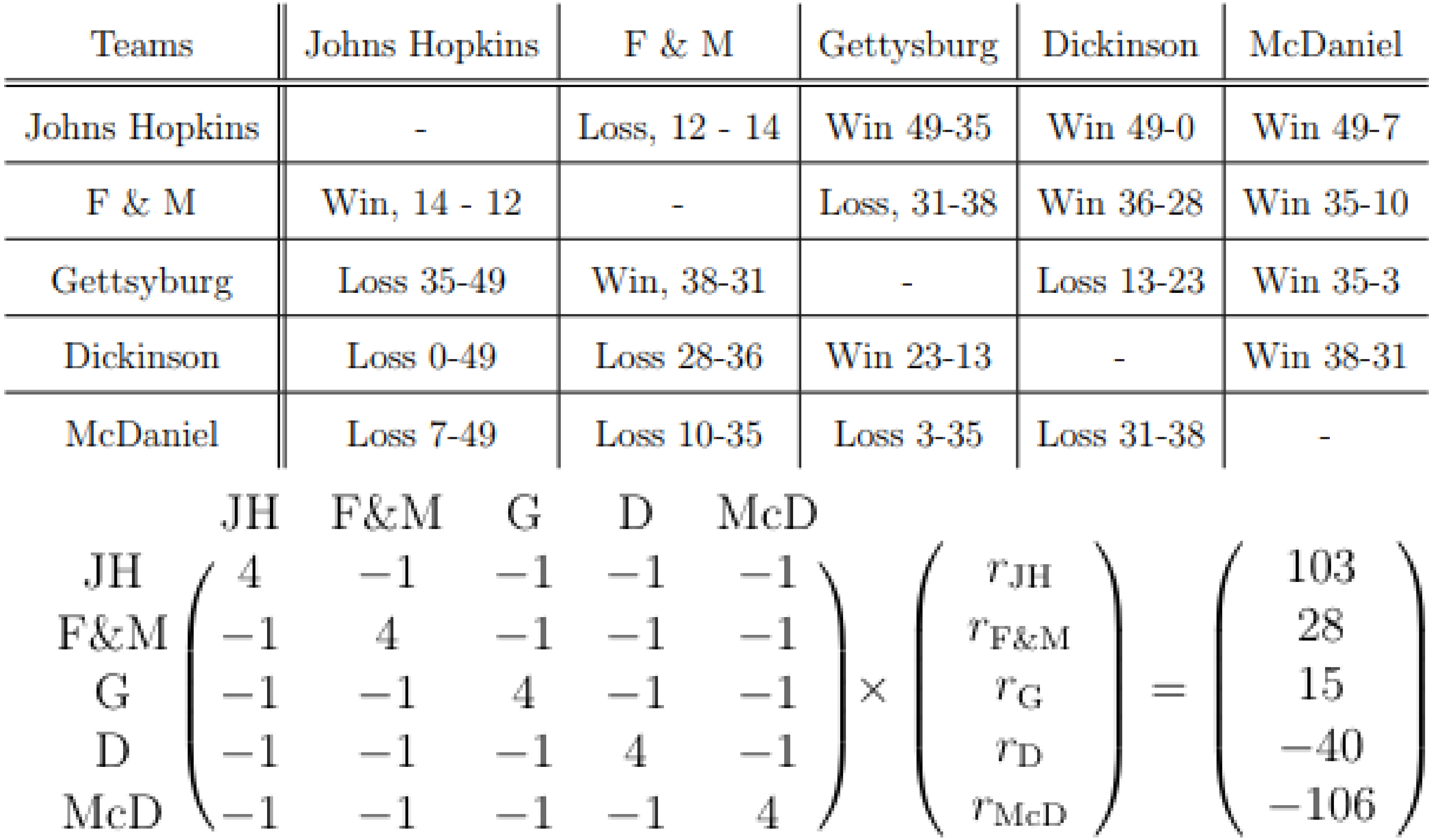
```
1 -2
```

Solving $Ax = b$ using $x = A^{-1}b$.

```
x <- solve(A)%*%b
print(x)
```

```
     [,1]
[1,]  0.0
[2,] -0.5
```

Matrix-Vector Multiplication Motivation



Teams	Johns Hopkins	F & M	Gettysburg	Dickinson	McDaniel
Johns Hopkins	-	Loss, 12 - 14	Win 49-35	Win 49-0	Win 49-7
F & M	Win, 14 - 12	-	Loss, 31-38	Win 36-28	Win 35-10
Gettsyburg	Loss 35-49	Win, 38-31	-	Loss 13-23	Win 35-3
Dickinson	Loss 0-49	Loss 28-36	Win 23-13	-	Win 38-31
McDaniel	Loss 7-49	Loss 10-35	Loss 3-35	Loss 31-38	-

$$\begin{array}{c|ccccc|} & \text{JH} & \text{F\&M} & \text{G} & \text{D} & \text{McD} \\ \text{JH} & 4 & -1 & -1 & -1 & -1 \\ \text{F\&M} & -1 & 4 & -1 & -1 & -1 \\ \text{G} & -1 & -1 & 4 & -1 & -1 \\ \text{D} & -1 & -1 & -1 & 4 & -1 \\ \text{McD} & -1 & -1 & -1 & -1 & 4 \end{array} \times \begin{pmatrix} r_{\text{JH}} \\ r_{\text{F\&M}} \\ r_{\text{G}} \\ r_{\text{D}} \\ r_{\text{McD}} \end{pmatrix} = \begin{pmatrix} 103 \\ 28 \\ 15 \\ -40 \\ -106 \end{pmatrix}$$

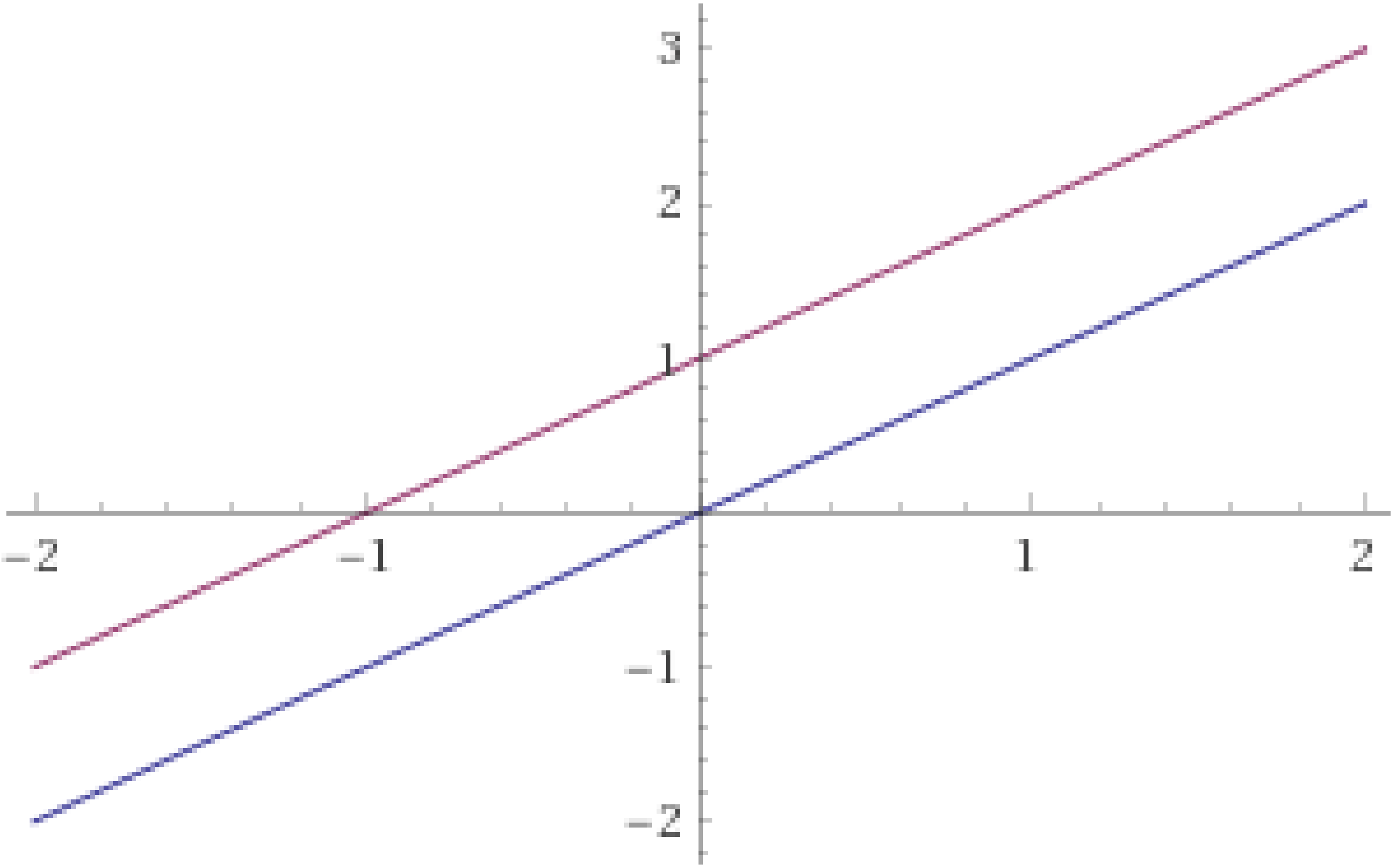
Matrix-Vector Multiplication Motivation

$$\begin{array}{c|ccccc} & \text{JH} & \text{F\&M} & \text{G} & \text{D} & \text{McD} \\ \hline \text{JH} & 4 & -1 & -1 & -1 & -1 \\ \text{F\&M} & -1 & 4 & -1 & -1 & -1 \\ \text{G} & -1 & -1 & 4 & -1 & -1 \\ \text{D} & -1 & -1 & -1 & 4 & -1 \\ \text{McD} & -1 & -1 & -1 & -1 & 4 \end{array} \times \begin{pmatrix} r_{\text{JH}} \\ r_{\text{F\&M}} \\ r_{\text{G}} \\ r_{\text{D}} \\ r_{\text{McD}} \end{pmatrix} = \begin{pmatrix} 103 \\ 28 \\ 15 \\ -40 \\ -106 \end{pmatrix}$$

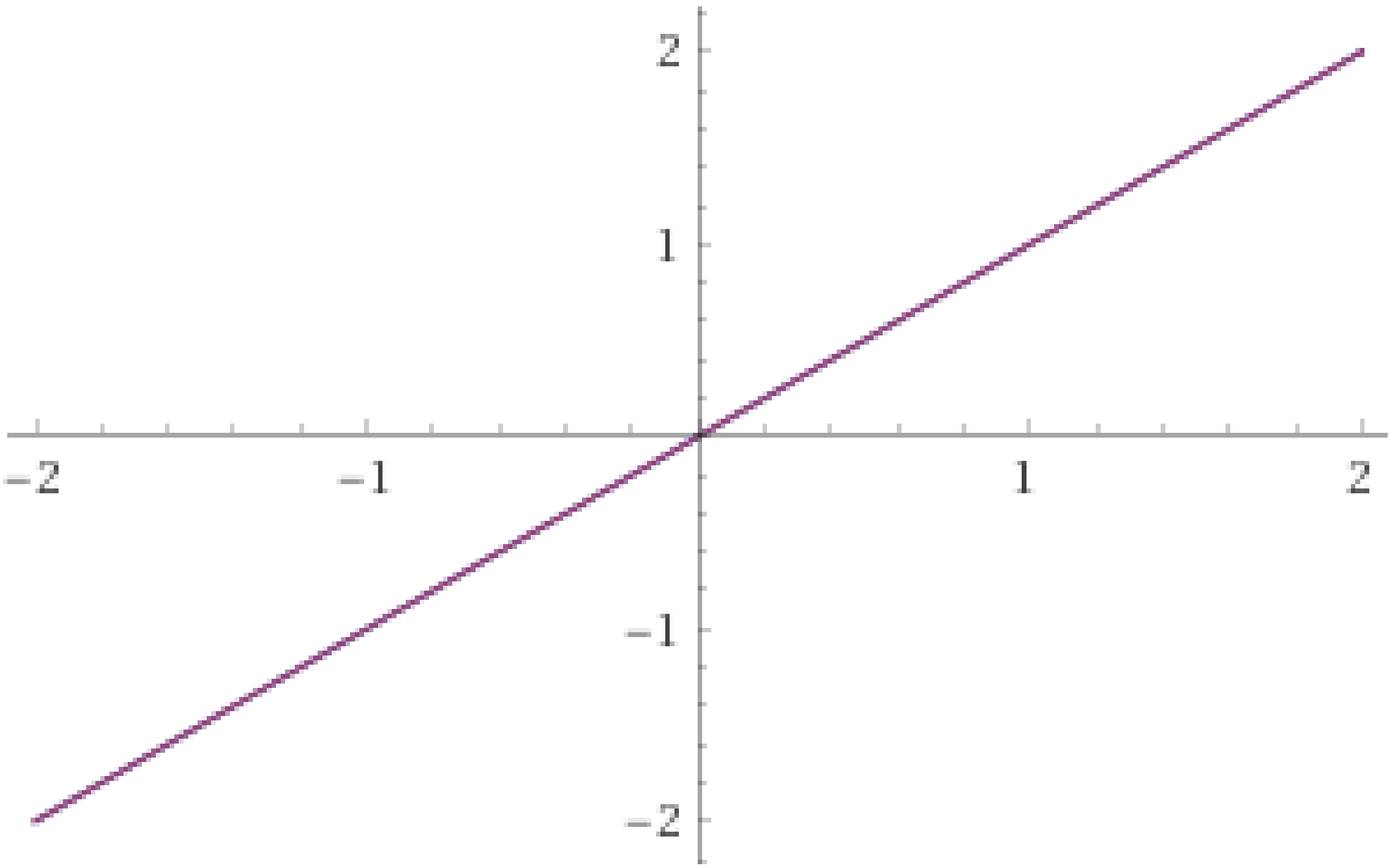
$$\text{inverse}\left(\begin{pmatrix} 4 & -1 & -1 & -1 & -1 & -1 \\ -1 & 4 & -1 & -1 & -1 & -1 \\ -1 & -1 & 4 & -1 & -1 & -1 \\ -1 & -1 & -1 & 4 & -1 & -1 \\ -1 & -1 & -1 & -1 & 4 & -1 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}\right)\left(\begin{pmatrix} 103 \\ 28 \\ 15 \\ -40 \\ -106 \\ \boxed{0} \end{pmatrix}\right) = \begin{pmatrix} 20.6 \\ 5.6 \\ 3 \\ -8 \\ -21.2 \\ 0 \end{pmatrix}$$

- Agregamos 0 al vector de respuesta para que la ultima columna no tenga ningun efecto sobre la respuesta de las variables x.
- El resultado de vemos en la ultima repuesta del vector x es la suma del vector, por causa del ejemplo la suma de puntos que gana un ejemplo termina siendo igual a la suma de puntos que pierde otro equipo.

Properties of Solutions to Matrix-Vector Equations - No Solutions



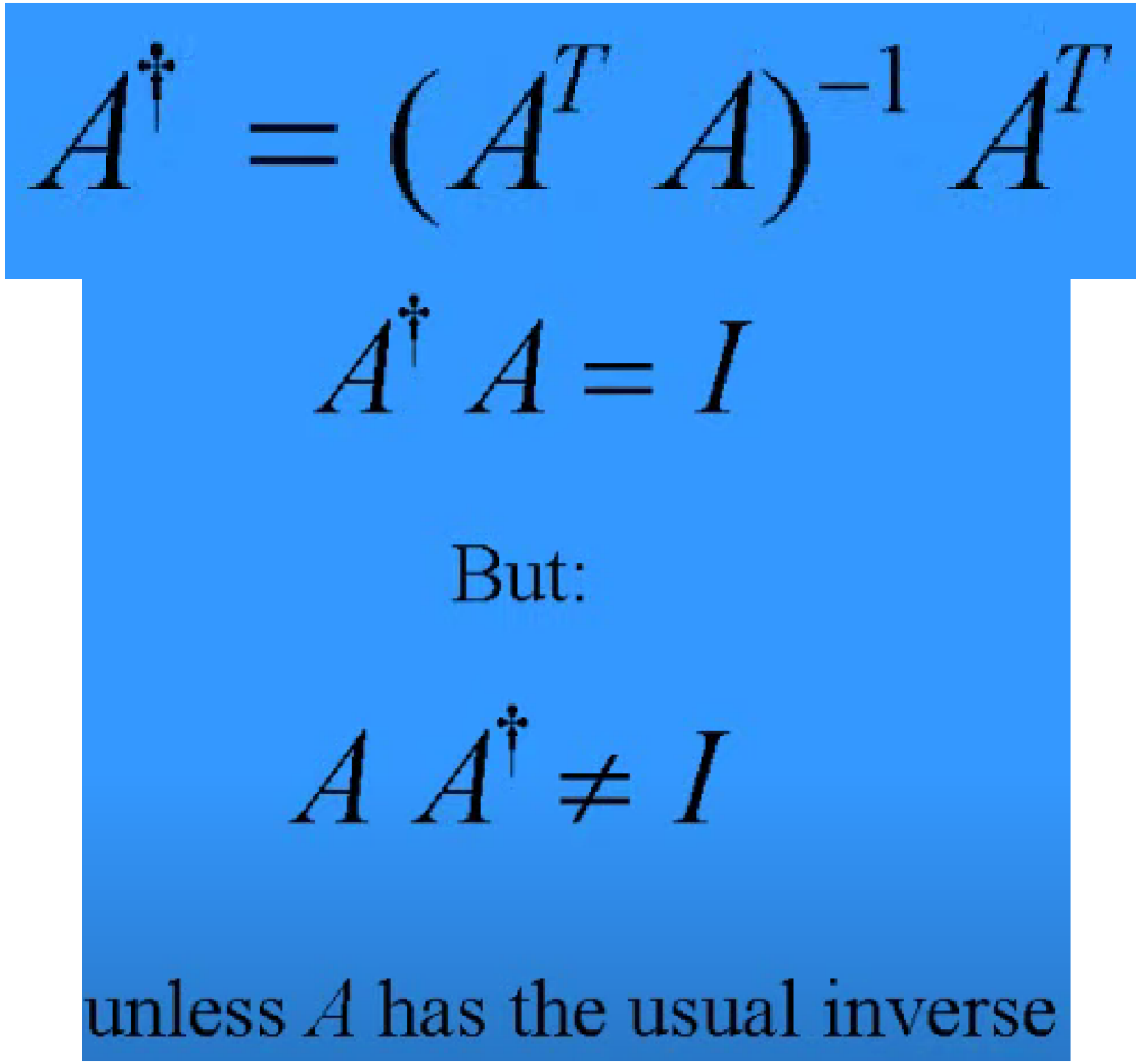
Properties of Solutions to Matrix-Vector Equations - Infinitely-Many Solutions



Some Options for Non-Square Matrices

- Row Reduction (By Hand, Difficult for Big Problems)
- Least Squares (If More Rows Than Columns - Used in Linear Regression)
- Singular Value Decomposition (If More Columns Than Rows - Used in Principal Component Analysis)
- Generalized or Pseudo-Inverse

Moore-Penrose Generalized Inverse



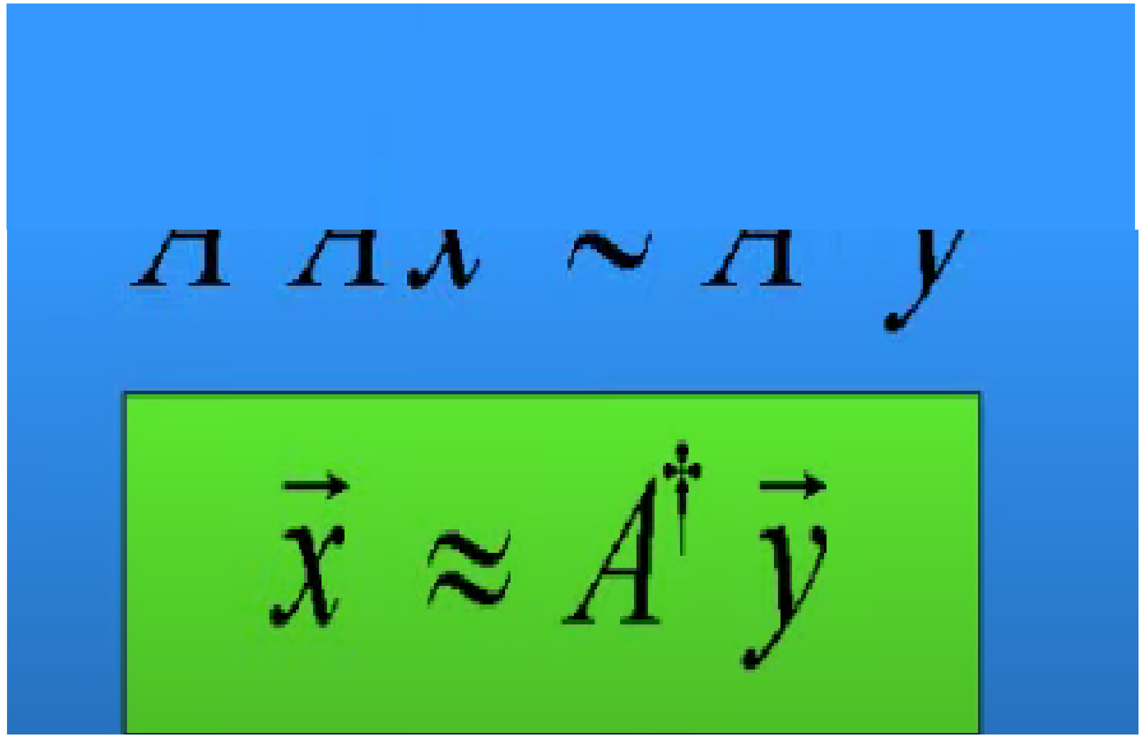
$$A^{\dagger} = (A^T A)^{-1} A^T$$

$$A^{\dagger} A = I$$

But:

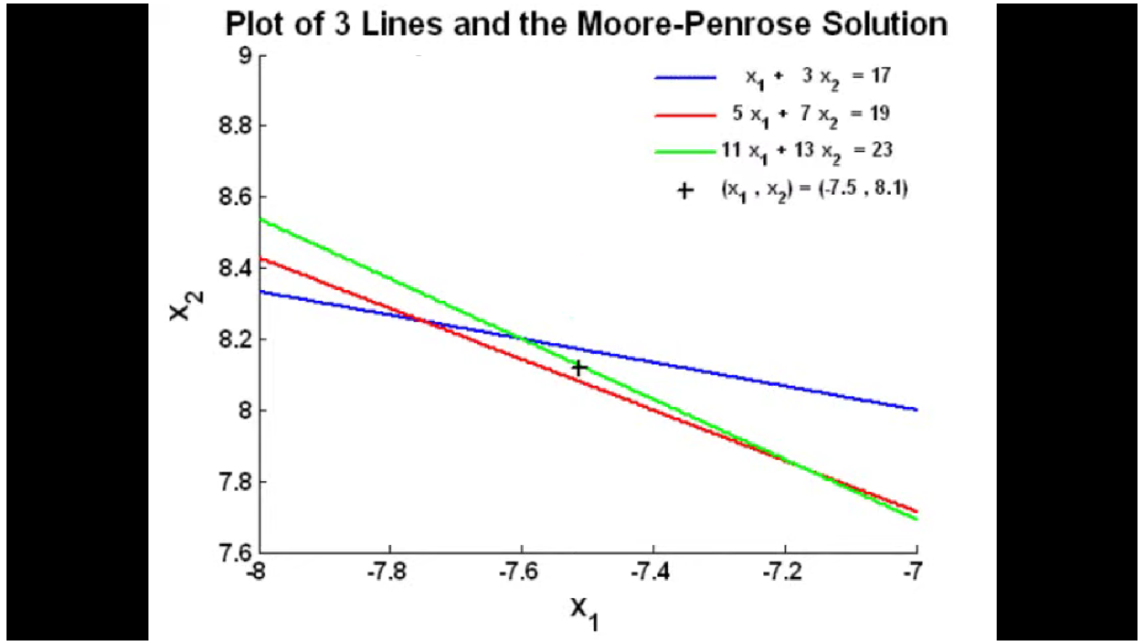
$$A A^{\dagger} \neq I$$

unless A has the usual inverse



$$A^T A x \approx A^T y$$

$$\vec{x} \approx A^{\dagger} \vec{y}$$



Plot of 3 Lines and the Moore-Penrose Solution

Moore-Penrose Generalized Inverse

```
library(MASS)
print(A)
```

```
     [,1] [,2]
[1,]    2    3
[2,]   -1    4
[3,]    1    7
```

```
ginv(A)
```

```
          [,1]        [,2]       [,3]
[1,] 0.3333333 -0.30303030 0.03030303
[2,] 0.0000000  0.09090909 0.09090909
```

```
ginv(A)%*%A
```

```
     [,1]          [,2]
[1,]    1 -1.110223e-16
[2,]    0  1.000000e+00
```

```
A%*%ginv(A)
```

```
           [,1]       [,2]      [,3]
[1,]  0.6666667 -0.3333333 0.3333333
[2,] -0.3333333  0.6666667 0.3333333
[3,]  0.3333333  0.3333333 0.6666667
```

Eigenvalues and Eigenvectors

LINEAR ALGEBRA FOR DATA SCIENCE IN R

Eric Eager
Data Scientist at Pro Football Focus

Eigenvalues and eigenvector

For a matrix A, the scalar λ is an eigenvalue of A, with associated eigenvector $v \neq 0$ if the following equation is true:

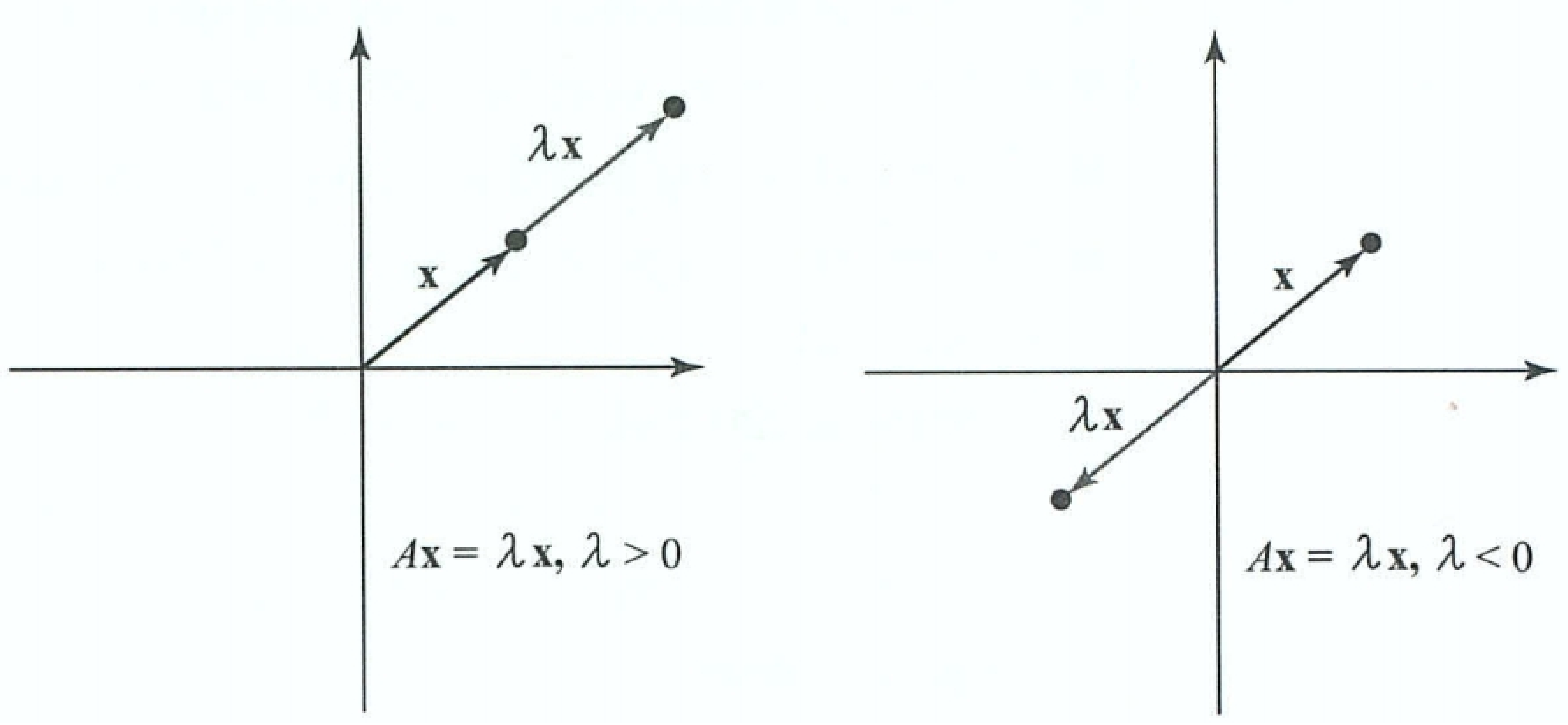
$$A\boldsymbol{v} = \lambda\boldsymbol{v}.$$

In other words:

The matrix multiplication $A\boldsymbol{v}$, a matrix-vector operation, produces the same vector as $\lambda\boldsymbol{v}$ a scalar multiplication acting on a vector.

This matrix does not have to be like the matrices in the last lecture.

Eigenvalues and eigenvector


$\lambda\mathbf{x}$
$\mathbf{x}$
$A\mathbf{x} = \lambda\mathbf{x},\ \lambda > 0$
$\mathbf{x}$
$\lambda\mathbf{x}$
$A\mathbf{x} = \lambda\mathbf{x},\ \lambda < 0$

Example

```
print(A)
```

```
     [,1] [,2]
[1,]    2    3
[2,]    0    1
```

Notice that $\lambda = 2$ is an eigenvalue of A with eigenvector $\vec{v} = (1, 0)^T$:

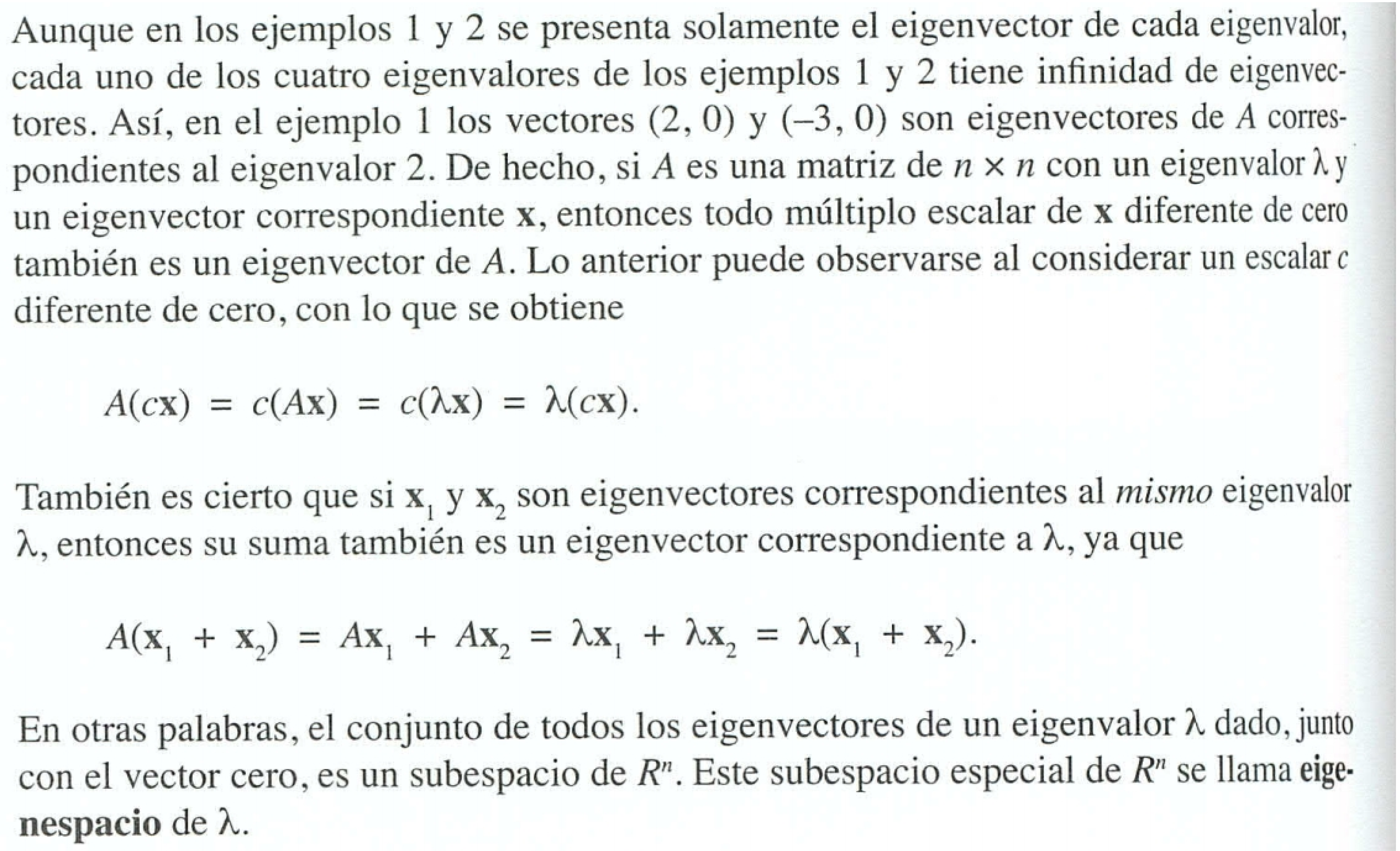
```
A%*%c(1,0)
```

```
     [,1]
[1,]    2
[2,]    0
```

```
2*c(1, 0)
```

```
2 0
```

Eigenespacio de λ



Aunque en los ejemplos 1 y 2 se presenta solamente el eigenvector de cada eigenvalor, cada uno de los cuatro eigenvalores de los ejemplos 1 y 2 tiene infinidad de eigenvectores. Así, en el ejemplo 1 los vectores (2, 0) y (–3, 0) son eigenvectores de A correspondientes al eigenvalor 2. De hecho, si A es una matriz de $n \times n$ con un eigenvalor λ y un eigenvector correspondiente $\mathbf{x}$, entonces todo múltiplo escalar de $\mathbf{x}$ diferente de cero también es un eigenvector de A. Lo anterior puede observarse al considerar un escalar c diferente de cero, con lo que se obtiene

$$A(c\mathbf{x}) = c(A\mathbf{x}) = c(\lambda\mathbf{x}) = \lambda(c\mathbf{x}).$$

También es cierto que si $\mathbf{x}_1$ y $\mathbf{x}_2$ son eigenvectores correspondientes al *mismo* eigenvalor λ, entonces su suma también es un eigenvector correspondiente a λ, ya que

$$A(\mathbf{x}_1 + \mathbf{x}_2) = A\mathbf{x}_1 + A\mathbf{x}_2 = \lambda\mathbf{x}_1 + \lambda\mathbf{x}_2 = \lambda(\mathbf{x}_1 + \mathbf{x}_2).$$

En otras palabras, el conjunto de todos los eigenvectores de un eigenvalor λ dado, junto con el vector cero, es un subespacio de R^n. Este subespacio especial de R^n se llama **eigenespacio** de λ.

Example, cont'd

Notice that $\lambda = 2$ is an eigenvalue of A with eigenvector $v = (1, 0)^T$ and $v = (4, 0)^T$:

```
A%*%c(1,0)
```

```
     [,1]
[1,]    2
[2,]    0
```

```
2*c(1, 0)
```

```
2 0
```

```
A%*%c(4,0)
```

```
     [,1]
[1,]    8
[2,]    0
```

```
2*c(4, 0)
```

```
8 0
```

Properties of Solutions to Eigenvalue/Eigenvector Problems

- An n by n matrix A has, up to multiplicity, n eigenvalues.
- Even if A is a matrix consisting entirely of real numbers, some (or all) of its eigenvalues could be complex numbers.
- All complex eigenvalues must come in conjugate pairs, though, like $1 + 2i$ and $1 - 2i$

```r
print(A)
```

```
     [,1] [,2] [,3]
[1,]   -1    2    4
[2,]    0    7   12
[3,]    0    0   -4
```

```r
eigen(A)
```

```
eigen() decomposition
$`values`
[1]  7 -4 -1

$vectors
          [,1]       [,2] [,3]
[1,] 0.2425356 -0.3789810    1
[2,] 0.9701425 -0.6821657    0
[3,] 0.0000000  0.6253186    0
```

```
print(A)
```

```
     [,1] [,2] [,3]
[1,]   -1    2    4
[2,]    0    7   12
[3,]    0    0   -4
```

Extracting eigenvalue and eigenvector information:

```
E <- eigen(A)
E$values[1]
E$vectors[, 1]
```

```
7
0.2425356 0.9701425 0.0000000
```

```
print(A)
```

```
     [,1] [,2]
[1,]    1    2
[2,]   -2   -1
```

```
eigen(A)
```

```
eigen() decomposition
$`values`
[1] 0+1.732051i 0-1.732051i

$vectors
                      [,1]                  [,2]
[1,]  0.3535534+0.6123724i  0.3535534-0.6123724i
[2,] -0.7071068+0.0000000i -0.7071068+0.0000000i
```

```
eigen(A)$values[1]*eigen(A)$values[2]
```

```
3+0i
```

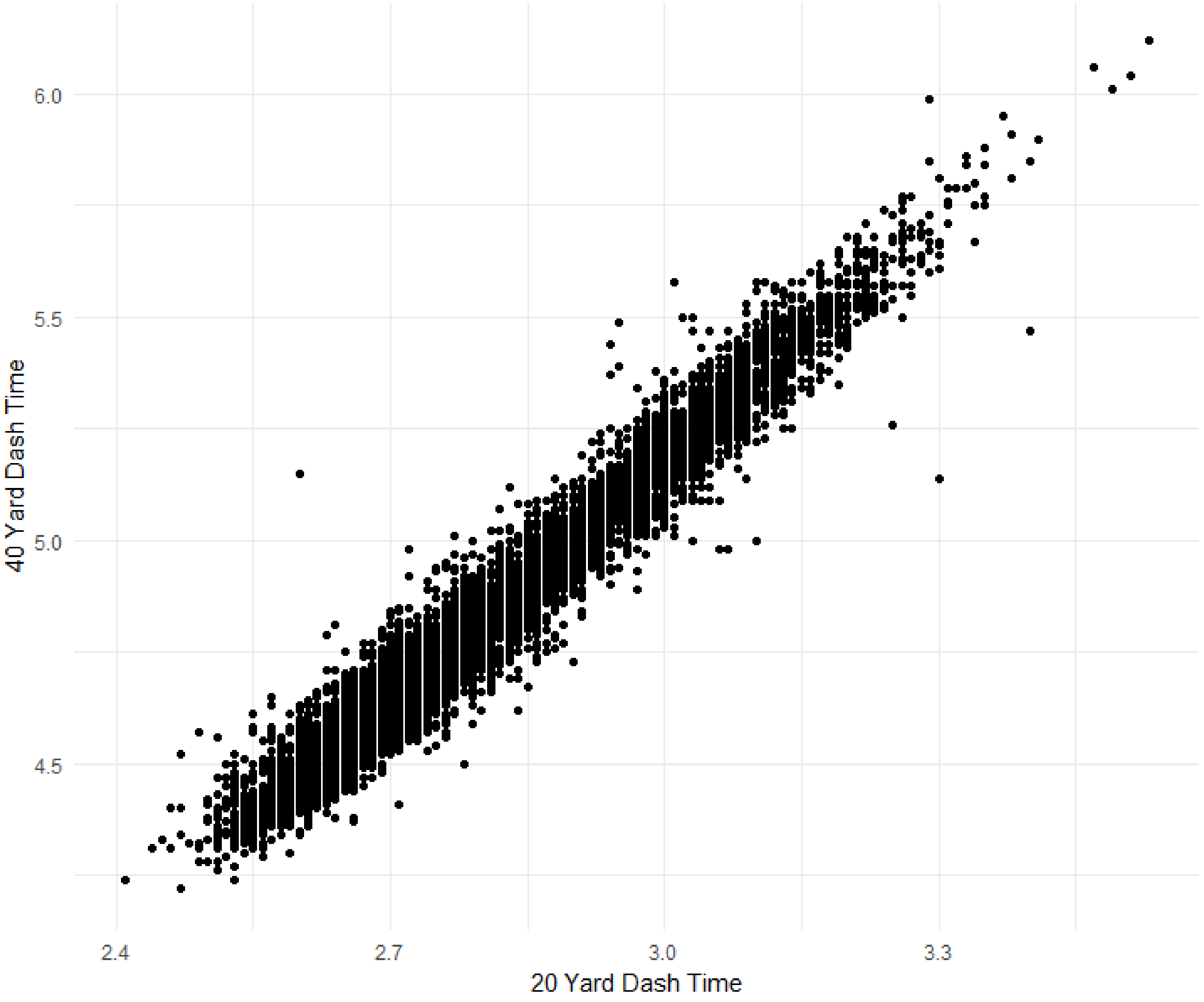
Principal Component Analysis

LINEAR ALGEBRA FOR DATA SCIENCE IN R

Eric Eager
Data Scientist at Pro Football Focus

datacamp

Big Data - Redundancy



Principal Component Analysis

- One of the more-useful methods from applied linear algebra
- Non-parametric way of extracting meaningful information from confusing data sets
- Uncovers hidden, low-dimensional structures that underlie your data
- These structures are more-easily visualized and are often interpretable to content experts

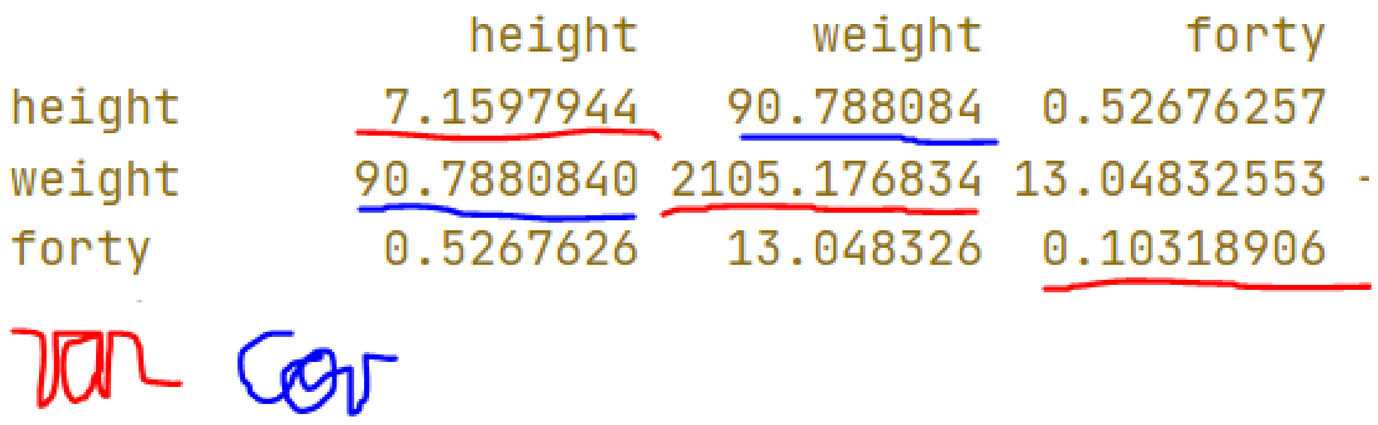
Theory

```
A <- scale(A)
```

$$\frac{A^T A}{n-1}$$

```
# Subtract the mean of each column
A[, 1] <- A[, 1] - mean(A[, 1])
A[, 2] <- A[, 2] - mean(A[, 2])
A[, 3] <- A[, 3] - mean(A[, 3])
A[, 4] <- A[, 4] - mean(A[, 4])
A[, 5] <- A[, 5] - mean(A[, 5])
A[, 6] <- A[, 6] - mean(A[, 6])
A[, 7] <- A[, 7] - mean(A[, 7])
A[, 8] <- A[, 8] - mean(A[, 8])
```

```
B <- t(A)%*%A/(nrow(A) - 1)
```



```
           height      weight       forty
height  7.1597944   90.788084  0.52676257
weight 90.7880840 2105.176834 13.04832553
forty   0.5267626   13.048326  0.10318906
```

var cor

PCA

- The eigenvalues $\lambda_1, \lambda_2, ..\lambda_n$ of $\frac{A^TA}{n-1}$ are real, and their corresponding eigenvectors are orthogonal, or point in distinct directions.
- The **total variance** of the data set is the sum of the eigenvalues of $\frac{A^TA}{n-1}$.
- These eigenvectors $v_1, v_2, ..., v_n$ are called the principal components of the data set in the matrix A.
- The direction that v_j points in can explain λ_j of the total variance in the data set. If λ_j, or a subset of $\lambda_1, \lambda_2, ..\lambda_n$ explain a significant amount of the total variance, there is an opportunity for **dimension reduction**.

Example

```
eigen(t(A)%*%A/(nrow(A) - 1))
```

```
eigen() decomposition
$`values`
[1] 12.5  0.0


$vectors
          [,1]       [,2]
[1,] 0.4472136 -0.8944272
[2,] 0.8944272  0.4472136
```

NFL Combine Data

```
head(select(combine, height:shuttle))
```

```
  height weight forty vertical bench broad_jump three_cone shuttle
1     71    192  4.38     35.0    14        127       6.71    3.98
2     73    298  5.34     26.5    27         99       7.81    4.71
3     77    256  4.67     31.0    17        113       7.34    4.38
4     74    198  4.34     41.0    16        131       6.56    4.03
5     76    257  4.87     30.0    20        118       7.12    4.23
6     78    262  4.60     38.5    18        128       7.53    4.48
```

NFL Combine Data

```r
prcomp(A)
```

```
Standard deviations (1, .., p=8):
[1] 46.7720885  6.6356959  4.7108443  2.2950226  1.6430770  0.2513368  0.1216908  0.1143365

Rotation (n x k) = (8 x 8):
                    PC1          PC2           PC3           PC4          PC5          PC6           PC7           PC8
height      0.042047079 -0.061885367  0.1454490039 -0.1040556410 -0.980792060  0.020679696 -6.155636e-03  0.0008055445
weight      0.980711529 -0.130912788  0.1270100265  0.0193388930  0.066908382 -0.008423587  6.988341e-04  0.0036087841
forty       0.006112061  0.012525260  0.0025260713 -0.0021291637  0.004096693  0.152469227 -2.539868e-01 -0.9549983725
vertical   -0.062926466 -0.333556369  0.0398922845  0.9366594549 -0.074901137  0.012214516  7.045063e-03 -0.0070051256
bench       0.088291423 -0.313533433 -0.9363461471 -0.0745692157 -0.107188391  0.009167322 -8.604309e-05 -0.0048308793
broad_jump -0.156742686 -0.876925849  0.2904565302 -0.3252903706  0.126494599  0.013753112 -2.187651e-03 -0.0076907609
three_cone  0.007468520  0.014691994  0.0009057581  0.0003320888  0.020902644  0.894560357 -3.743559e-01  0.2427137770
shuttle     0.004518826  0.009863931  0.0023111814 -0.0094052914  0.004010629  0.419039274  8.917710e-01 -0.1700673446
```

NFL Combine Data

```r
summary(prcomp(A))
```

```
Importance of components:
                           PC1     PC2     PC3     PC4     PC5     PC6     PC7     PC8
Standard deviation     46.7721 6.63570 4.71084 2.29502 1.64308 0.25134 0.12169 0.11434
Proportion of Variance  0.9672 0.01947 0.00981 0.00233 0.00119 0.00003 0.00001 0.00001
Cumulative Proportion   0.9672 0.98663 0.99644 0.99877 0.99996 0.99999 0.99999 1.00000
```

```r
head(prcomp(A)$x[, 1:2])
```

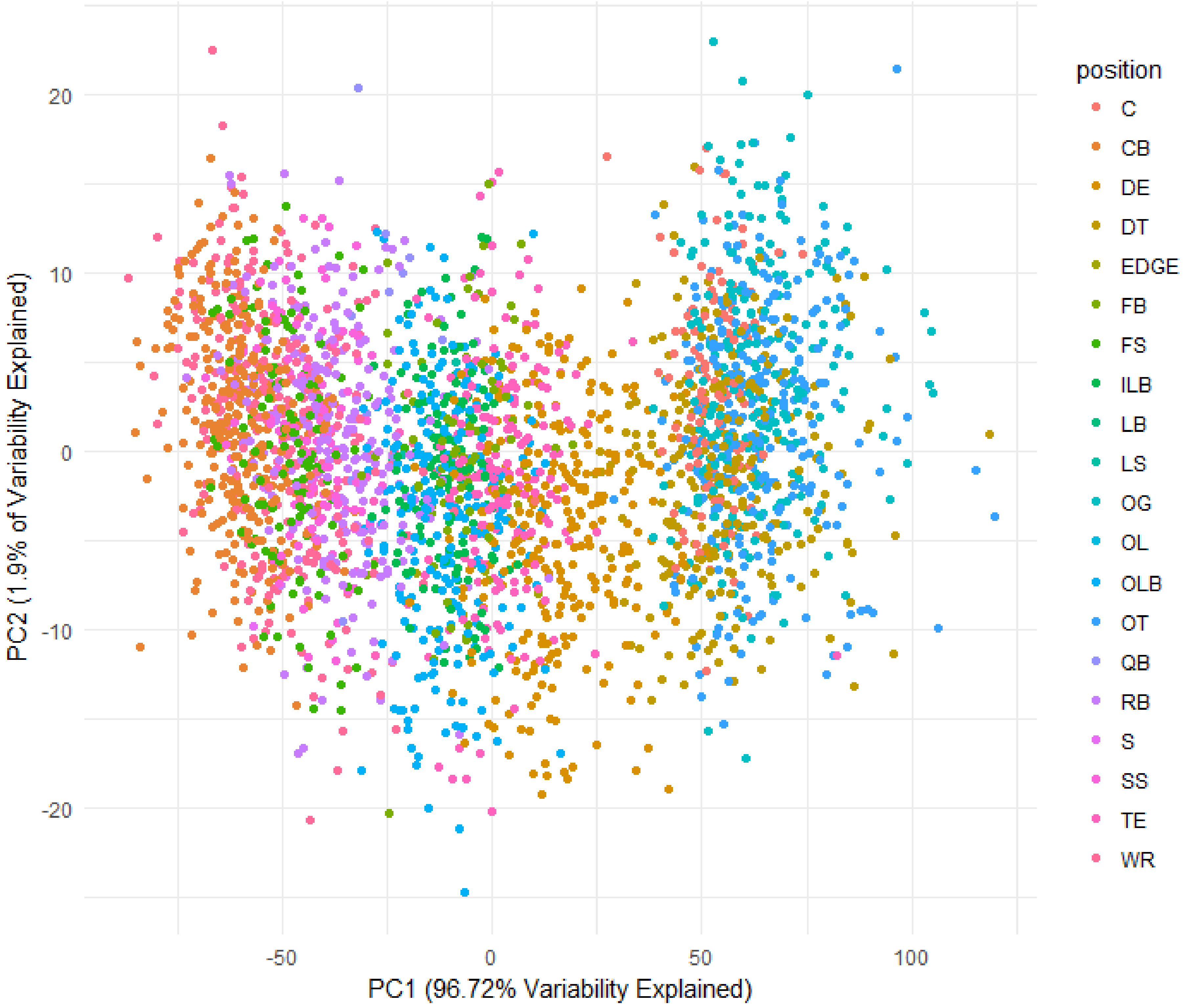
```
            PC1        PC2
[1,] -62.005067  -2.654645
[2,]  48.123290   6.693433
[3,]   3.732016   1.283046
[4,] -56.823742  -9.764098
[5,]   4.213670  -3.779862
[6,]   6.924978 -15.530509
```

```r
head(cbind(combine[, 1:4], prcomp(A)$x[, 1:2]))
```

```
             player position        school year        PC1        PC2
1   Jaire Alexander       CB    Louisville 2018 -62.005067  -2.654645
2       Brian Allen        C Michigan St. 2018  48.123290   6.693433
3      Mark Andrews       TE      Oklahoma 2018   3.732016   1.283046
4         Troy Apke        S     Penn St. 2018 -56.823742  -9.764098
5 Dorance Armstrong     EDGE        Kansas 2018   4.213670  -3.779862
6         Ade Aruna       DE        Tulane 2018   6.924978 -15.530509
```

Things to Do After PCA

- Data wrangling/quality control
- Data visualization
- Unsupervised learning (clustering)
- Supervised learning
- Much more!


PC2 (1.9% of Variability Explained)
PC1 (96.72% Variability Explained)
position
● C
● CB
● DE
● DT
● EDGE
● FB
● FS
● ILB
● LB
● LS
● OG
● OL
● OLB
● OT
● QB
● RB
● S
● SS
● TE
● WR