University of Toronto
CSC**343**

CSC343 Assignment 3

Question 1)
(a)
$S^+ = QSTUVX$, violates BCNF as R is not in RHS.
$R^+ = QRSTUVWX$, R is superkey
$V^+ = QUV$, violates BCNF as V clearly not superkey
$S^+ = QSTUVX$, violates BCNF, we already did this calculation above.

So, $S \rightarrow U$, $V \rightarrow QU$ and $S \rightarrow TVX$ violate BCNF

(b)
$R1 = QRSTUVWX$
We found in part a that $S \rightarrow U$ violates BCNF. Given $S^+ = QSTUVX$, we use this violating functional dependency (FD) to break R1 into $R2 = RSW$ and $R3 = QSTUVX$.

Now lets project the original FDs on R2 to see if there is any FD in R2 that violates BCNF (to see if we should break even further). We write "skip" when a subset of the attributes in question is already a superkey when taking the closure. We write "nothing" when there is no non-trivial FD that can be derived from the closure of the left hand side.

R	S	W	closure	FDs
✓			$R^+ = QRSTUVWX$	$R \rightarrow SW$, is superkey so doesnt violate BCNF
	✓		$S^+ = QSTUVX$	nothing
		✓	$W^+ = W$	nothing
✓	✓		skip	skip
✓		✓	skip	skip
	✓	✓	$SW^+ = QSTUVWX$	nothing

So, $R2 = RSW$, since no FD violates BCNF.

Now, lets do the same for $R3 = QSTUVX$.

Q	S	T	U	V	X	closure	FDs
✓						$Q^+ = Q$	nothing
	✓					$S^+ = QSTUVX$	$S \rightarrow QTUVX$, superkey of R3 so doesnt violate BCNF
		✓				$T^+ = T$	nothing
			✓			$U^+ = U$	nothing
				✓		$V^+ = QUV$	$V \rightarrow QU$ violates BCNF, abort the projection

$R3 = QSTUVX$ with $V \rightarrow QU$ violating BCNF, so break into $R4 = QUV$ and $R5 = STVX$.

As shown by table below, there is no FD in R4 violating BCNF, so no need to break R4 further.

Q	U	V	closure	FDs
✓			$Q^+ = Q$	nothing
	✓		$U^+ = U$	nothing
		✓	$V^+ = QUV$	$V \to QU$, is superkey
✓	✓		$QU^+ = QU$	nothing
✓		✓	skip	skip
	✓	✓	skip	skip

As shown by the table below, there is no FD in R5 either violating BCNF, so no need to break R5 further.

V	S	T	X	closure	FDs
✓				$V^+ = QUV$	nothing
	✓			$S^+ = QSTUVX$	$S \to TVX$, is superkey
		✓		$T^+ = T$	nothing
			✓	$X^+ = X$	nothing
✓	✓			skip	skip
✓		✓		$VT^+ = QTUV$	nothing
✓			✓	$VX^+ = QUVX$	nothing
	✓	✓		skip	skip
	✓		✓	skip	skip
		✓	✓	$TX^+ = TX$	nothing
✓	✓	✓		skip	skip
✓	✓		✓	skip	skip
✓		✓	✓	$VTX^+ = QTUVX$	nothing
	✓	✓	✓	skip	skip

So the final breakdown is $R2 = RSW$, $R4 = QUV$ and $R5 = STVX$. The FD for each relation is clear from just reading off their respective projection tables. For R2: $R \to SW$, for R4: $V \to QU$, for R5: $S \to TVX$

(c) **No**, we can formally project functional dependencies into our new relations to figure this out. When we do so (see q1b), we observe that the dependency $S \to U$ is no longer preserved in any of the tables. In fact, no table has both S and U as attributes.

(d) Say that a = (q,r,s,t,u,v,w,x) appears in the join of R2, R4, R5. We will prove that it must appear in the original R0. We do this using the Chase test. If a appears in the join, the original must have a row with (r,s,w), a row with (s,t,v,x), and a row with (q,u,v). This is because projecting attributes to R2, R4, R5 will not mix rows for us, so they must already exist in those groups. So, R0 must have the following structure.

Q	R	S	T	U	V	W	X
1	r	s	2	3	4	w	5
6	7	s	t	8	v	9	x
q	10	11	12	u	v	13	14

Now, to make the first row satisfy $S \to TVX$, it must be adjusted as shown below. This is because row 1 and 2 both have the same value for S, so row 1 must also copy row 2's values for T,V,X.

Q	R	S	T	U	V	W	X
1	r	s	t	3	v	w	x
6	7	s	t	8	v	9	x
q	10	11	12	u	v	13	14

Finally, to satisfy $V \to QU$, we must adjust the first row as shown below. This is because row 1 and 3 both have the same value for V, so row 1 must copy over row 3's values for Q,U.

Q	R	S	T	U	V	W	X
q	r	s	t	u	v	w	x
6	7	s	t	8	v	9	x
q	10	11	12	u	v	13	14

Now, the first row is (q,r,s,t,u,v,w,x), which means if this row occurs in the join, it must have appeared in the original R1. And thus, the join is lossless, it doesnt add any new rows.

Question 2)
(a)
Step 1: Split the RHSs to get our initial set of FDs, S1:
(i) $CDH \to F$
(ii) $G \to D$
(iii) $G \to H$
(iv) $FG \to C$
(v) $FG \to D$
(vi) $FG \to E$
(vii) $H \to C$
(viii) $H \to E$
(ix) $H \to G$
(x) $F \to C$
(xi) $F \to D$

Step 2: For each FD, try to reduce the LHS:
(i) The only way to get F is from this FD, so we cannot reduce it.
(ii) Only one attribute on LHS, so can't reduce
(iii) Can't reduce, same reason as ii
(iv) $F^+ = FCD$, so reduce $FG \to C$ to $F \to C$.
(v) $F^+ = FCD$, so reduce $FG \to D$ to $F \to D$.
(vi) $G^+ = GDHCEF$ so reduce $FG \to E$ to $G \to E$
(vii) Can't reduce, same reason as ii
(viii) Can't reduce, same reason as ii
(ix) Can't reduce, same reason as ii
(x) Can't reduce, same reason as ii
(xi) Can't reduce, same reason as ii

Our new set of FDs, call it S2, in alphabetical ordering is:
(1) $CDH \to F$
(2) $F \to C$
(3) $F \to C$
(4) $F \to D$
(5) $F \to D$
(6) $G \to D$
(7) $G \to E$
(8) $G \to H$
(9) $H \to C$
(10) $H \to E$
(11) $H \to G$

Step 3: try to eliminate each FD.

FD	Exclude these from $S2$ when computing closure	Closure	Decision
1	1	There's no way to get F without this FD	keep
2	2	$F^+ = FCD$	discard
3	2, 3	$F^+ = FD$	keep
4	2, 4	$F^+ = FCD$	discard
5	2, 4, 5	$F^+ = FC$	keep
6	2, 4, 6	$G^+ = GHEC$	keep
7	2, 4, 7	$G^+ = GDHCE$	discard
8	2, 4, 7, 8	$G^+ = GD$	keep
9	2, 4, 7, 9	$H^+ = HEGD$	keep
10	2, 4, 7, 10	$H^+ = HCGDF$	keep
11	2, 4, 7, 11	$H^+ = HCE$	keep

Our final set of FDs (combining RHS whenever LHS is same) are
$CDH \rightarrow F$
$F \rightarrow CD$
$G \rightarrow DH$
$H \rightarrow CEG$

(b) A, B are not in LHS or RHS of any FD, so they must appear in any key. E appears in RHS but never in LHS, so it is never in any key. So, we only need to check for T = {C, D, F, G, H} - which minimal subset of T gives a closure that includes all of the attributes in T. This is easy. $H^+ = HCEGDF$, which includes all of T. $G^+ = GDHCEF$, which also includes all of T. So, H and G form a key each when combined with A,B. Now we must see if there is a way to cover all of T with just F, C, D. Well, $FCD^+ = FCD$, so even if we include all of them, we don't get a key as we cant cover G or H.
So, the only 2 keys are:
ABG
ABH

(c)
We follow 3NF algorithm:
1. Make sure FDs are a minimal basis. We already did this in part a.

2. Make a relation from every FD.
$CDH \rightarrow F$ gives $R1 = CDFH$
$F \rightarrow CD$ gives $Rtemp = CDF$
$G \rightarrow DH$ gives $R2 = DGH$
$H \rightarrow CEG$ gives $R3 = CEGH$
But $Rtemp$ is a subset of R1, so we remove it.

3. None of these relations R1 to R3 can be superkeys because they do not give A, B. So, we must make a relation from a key, say ABG. Now, our final 3NF synthesis relations are:
$R1 = CDFH$
$R2 = DGH$
$R3 = CEGH$
$R4 = ABG$

(d) It's known that 3NF can allow FDs with a non-superkey on the LHS. This allows redundancy, and thus anomalies. In our case, consider R1. After projecting the minimal basis FDs on this relation, we have $F^+ = FCD$, which gives $F \rightarrow CD$. This is not a superkey as it does not include H. So, our decomposition does allow redundancy.